

miostat

Статистика для MapInfo

Назначение программы

Программа miStat.mbx предназначена для выполнения первичного статистического анализа данных представленных в полях таблиц проекта MapInfo. Назначение такого анализа, не выходя из среды MapInfo, получить общее представление о данных, о связывающих их закономерностях в наиболее общем и наглядном виде, позволяющем сформулировать некоторые качественные утверждения, строгую проверку которых уже выполнять в специализированных статистических пакетах, таких как SPSS, STATISTICA, SAS, R и т.п.

Предлагаемое руководство содержит сведения о применении программного средства miStat и порядка работы с ним¹. Программа работает в среде MapInfo версии 10 и выше². При работе используются следующие библиотеки: ALGLIB® (v. 3.8.2), ZedGraph.dll (v. 5.1.5), miStat.dll, btSt.dll.

¹ Далее в тексте нет пояснений к определяемым статистикам и нет определений используемых терминов. Предполагается, что при необходимости пользователю несложно будет получить эту информацию из литературы или из интернета.

² Среда тестирования: Win7-64, .NET Framework (v. 3.5), MapBasic 11.5, MapInfo 11.5. Окончательная версия программы была скомпилирована в MapBasic 10.0.

Установка

Программа не требует специальной установки. Поддерживается развертывание XCOPY, то есть для установки нужно просто скопировать все файлы и папки приложения на жесткий диск без какой-либо регистрации. В данном случае нужно скопировать папку miStat со всеми входящими библиотеками и вложенной папкой mi_Data. После запуска программы³ (файл miStat.mbx) в окно MapInfo будет добавлена новая панель инструментов (Рис. 1). Требуется наличие на компьютере .NET Framework (v. 3.5).

³ Меню **Программы/Запустить программу MapBasic...** или просто переместить файл программы в окно MapInfo с помощью мыши.

Панель инструментов программы

После старта программы в окне MapInfo появится новая плавающая инструментальная панель. Кнопки этой панели связаны с функциями, подробное описание которых приводится ниже.



Рис. 1



Анализ выборки

Определяется набор основных статистических показателей эмпирической выборки то, что определяется термином «описательная статистика». Анализируемые данные (выборка) относятся к одному полю одной таблицы⁴. Минимальная длина выборки 10 элементов. После старта данной функции будет открыта форма для выбора таблицы и поля в этой таблице (Рис. 2а). Для открытия списка полей таблицы необходимо выполнить двойной клик на имени таблицы (Рис. 2б). Кроме того, по необходимости, нужно установить переключатели «запись протокола» и «сохранение преобразованных данных».

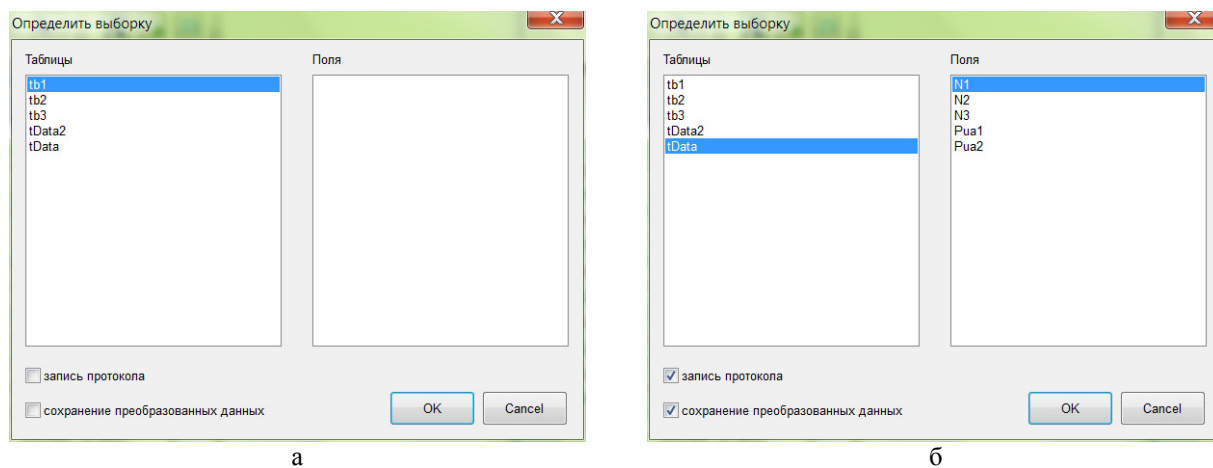


Рис. 2

В случае выбора переключателя «запись протокола» будет сформирован файл протокола, в который будет выводиться информация, определяемая данной процедурой. Файл будет записан в папку *mi_Data* с именем, образуемым по следующему правилу: *ггммдд_ччммсс_sd.html*⁵ (например, 140807_001436_sd.html). Если установлен второй переключатель, то для каждого выбранного преобразования, в базовой таблице будет создано новое поле, которое будет заполнено преобразованными данными. Имя поля образуется по правилу:

ИмяБазовогоПоля_КодПреобразования

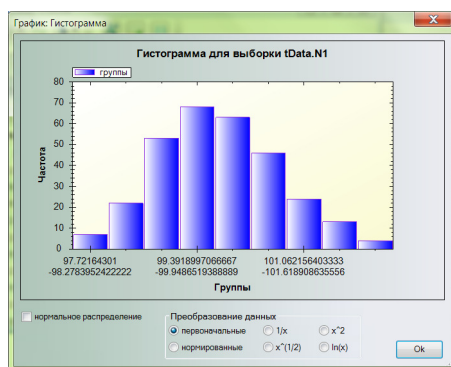
Так для базового поля X можно образовать следующие поля для преобразованных данных.

X_N	Центрирование и нормализация по СКО
X_1X	Преобразование «1/x»
X_SQR	Преобразование - корень квадратный из «x»
X_X2	Преобразование - «x» в степени 2
X_LN	Преобразование - «Ln(x)»

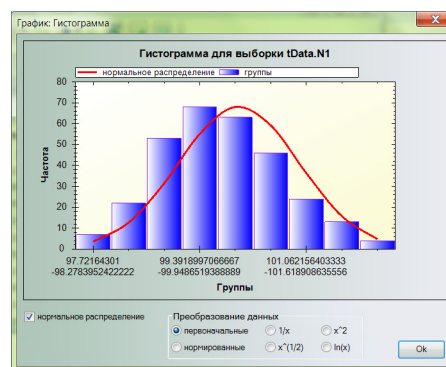
⁴ Таблица может содержать графические объекты, но может и не иметь связи с картой.

⁵ Не должно смущать, что файл протокола записывается в формате html. Его можно без проблем открыть практически в любом текстовом редакторе (MS Word, LibreOffice и т.д.), представить в необходимом виде и распечатать как обычный документ. При необходимости можно выполнить экспорт в нужный формат (doc, docx, pdf и т.д.).

На Рис. 6б показана таблица с добавленными полями преобразованных данных. После нажатия кнопки «ОК» (форма «Определить выборку») открывается новая форма (Рис. 3а) представляющая гистограмму для текущей выборки. Здесь же можно выбрать необходимые преобразования для выборки. Гистограмма автоматически перестраивается под новые данные. Если выбрать переключатель «нормальное распределение» то на гистограмму будет наложена кривая плотности нормального распределения. Кривая построена для математического ожидания и дисперсии, приравненных к соответствующим оценкам на текущей выборке (Рис. 3б).



а



б

Рис. 3

Аналогично и на преобразованных данных (Рис. 4). Цифра в скобках в заголовке гистограммы определяет код преобразования. Размеры формы можно менять по необходимости.

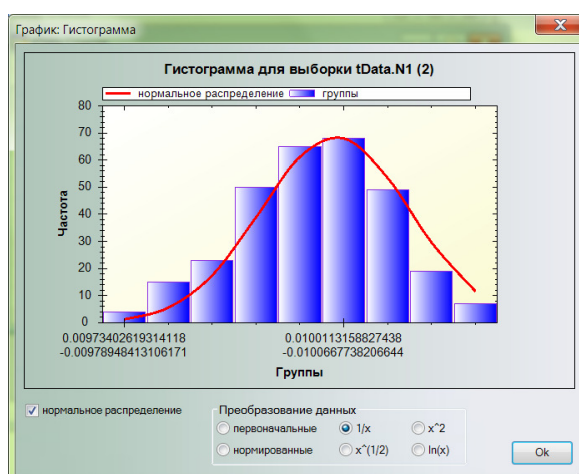
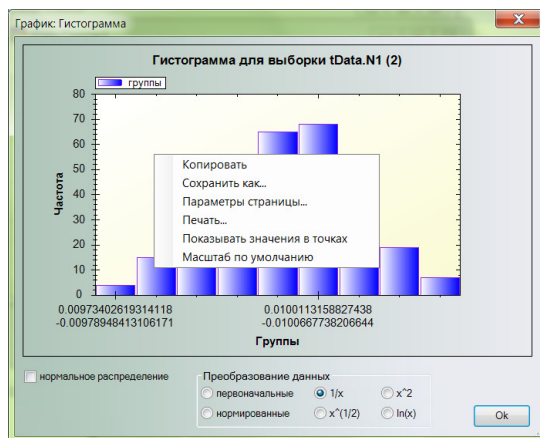


Рис. 4

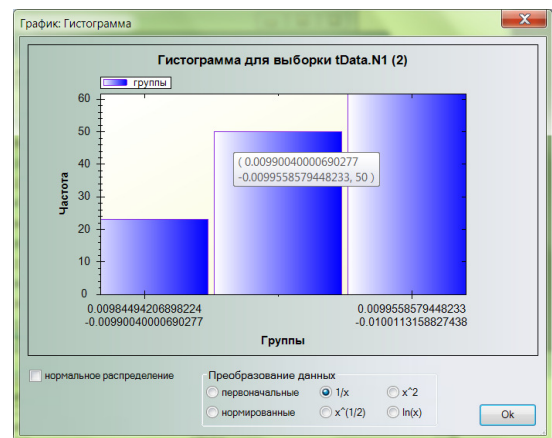
Для области гистограммы доступно контекстное меню, вызываемое по нажатию правой кнопки мыши (Рис. 5а). Кроме того можно увеличить отдельную часть графика если, нажав левую кнопку мыши, выделить интересующую область графика (Рис. 5б). Возврат к исходному виду выполняется выбором в контекстном меню пункта «Масштаб по умолчанию».

Независимо от установки записи информации в протокол, в окно «Сообщение» всегда выводится информация по базовому полю выбранной таблицы (Рис. 6а). В файл протокола также выводится информация и по полям с преобразованными данными. Некоторые из определяемых статистик предполагают, что распределение стремится к нормальному распределению. Прямого анализа на нормальное распределение не выполняется. Однако об этом можно судить по виду гистограммы и ее соответствия кривой плотности нормального распределения (хотя для выборок небольшого объема это

проблематично). Косвенно также по значению $SKO/cp. oш.$, которое в случае нормального распределения приблизительно равно 1.25. Кроме того в какой то мере об этом позволяют судить значения асимметрии (A) и эксцесса (E). Считается, что отклонение от нуля этих величин не противоречит гипотезе о нормальном распределении, если выполняются неравенства $A \leq 3 * SQR(Da)$ и $|E| \leq 5 * SQR(De)$.



а



б

Рис. 5

Сообщение

ОПИСАТЕЛЬНАЯ СТАТИСТИКА:

Таблица	tData
Поле	N1
Число элементов	300
Сумма	30001.4328515
Макс. значение	102.7324131
Мин. значение	97.72164301
Среднее значение	100.004776171667
SKO среднего (0.95)	0.054403950331624
Границы дисперсии (0.95)	99.9150115907476, 100.094540752586
Дисперсия	0.887935464857206
Границы дисперсии (0.95)	0.76117263321369, 1.04940960021228
SKO	0.942303276475895
Среднее отклонение	0.766551097577777
SKO/Среднее отклонение	1.22927653414557
Медиана	99.9470765
Процентиль (0.00)	97.72164301
Процентиль (0.25)	99.3411682925
Процентиль (0.50)	99.9470765
Процентиль (0.75)	100.61611335
Процентиль (1.00)	102.7324131
Асимметрия	0.232263778185779
3*SQR(Da)	0.424264068711929
Эксцесс	-0.20130640926736
5*SQR(De)	1.41421356237309

а

tData Список

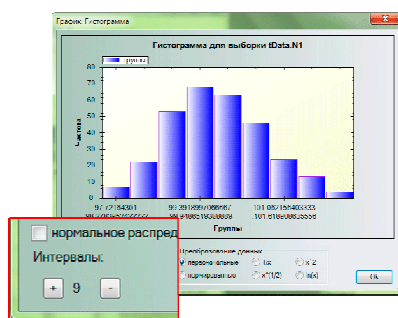
N1	N2	N3	Pua1	Pua2	N1_N	N1_1X	N1_SQR
101.552	100.069	97.4287	0	0	1.64246	0.00984713	10.0773
98.8998	100.512	100.645	1	2	-1.17263	0.0101112	9.94484
99.1291	99.4608	99.9168	0	1	-0.929308	0.0100879	9.95636
99.8874	100.351	99.0513	3	0	-0.124536	0.0100113	9.99437
99.3736	100.204	99.4002	0	0	-0.669773	0.010063	9.96863
99.9406	101.417	99.2865	1	0	-0.0681433	0.0100059	9.99703
99.7831	100.729	100.832	1	2	-0.235268	0.0100217	9.98915
100.641	99.7972	100.11	1	0	0.674818	0.00993634	10.032
98.8189	100.575	101.696	0	0	-1.25851	0.0101195	9.94077

б

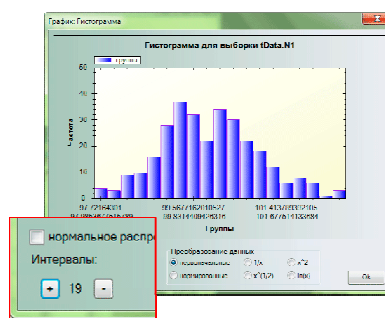
Рис. 6

Добавлено 29.08.2014:

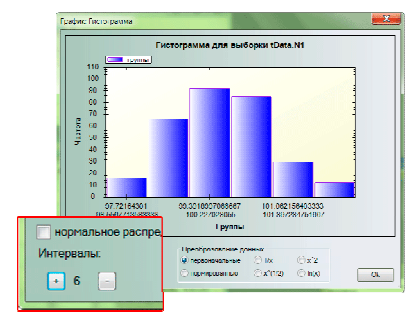
Число интервалов для гистограммы выбирается автоматически (Рис. 7а), но в некоторых случаях возникает желание изменить число интервалов. С этой целью на форму «Гистограмма» добавлены две кнопки $+$ и $-$ и текстовое поле, показывающее текущее значение числа интервалов. На Рис. 7б, 7в число интервалов изменено.



а



б



в

Рис. 7

Анализ выборки

16 августа 2014 г., 19:45:20

Базовая выборка из таблицы tData поле N1

ОПИСАТЕЛЬНАЯ СТАТИСТИКА:

Таблица_____tData
Поле_____N1
Число элементов_____300
Сумма_____30001.4328515
Макс. значение_____102.7324131
Мин. значение_____97.72164301
Среднее значение_____100.004776171667
СКО среднего_____0.0544039050331624
Границы среднего (0.95)_____99.9150115907476, 100.094540752586
Дисперсия_____0.887935464857206
Границы дисперсии (0.95)_____0.76117263321369, 1.04940960021228
СКО_____0.942303276475895
Среднее отклонение_____0.766551097577777
СКО/Среднее отклонение_____1.22927653414557
Медиана_____99.9470765
Процентиль (0.00)_____97.72164301
Процентиль (0.25)_____99.3411682925
Процентиль (0.50)_____99.9470765
Процентиль (0.75)_____100.61611335
Процентиль (1.00)_____102.7324131
Ассиметрия_____0.232263778185779
3*SQR(Da)_____0.424264068711929
Эксцесс_____ -0.20130640926736
5*SQR(De)_____1.41421356237309

ОПИСАТЕЛЬНАЯ СТАТИСТИКА для преобразованных данных:

Тип преобразования_____нормирование
Число элементов_____300
Сумма_____1.03823616370846E-11
Макс. значение_____2.89464867248941
Мин. значение_____ -2.42292817892483
Среднее значение_____3.46078721236154E-14
СКО среднего_____0.0577350269189626
Границы среднего (0.95)_____ -0.0952608179976938, 0.095260817997763
Дисперсия_____1
Границы дисперсии (0.95)_____0.857238688327534, 1.18185345866441
СКО_____1
Среднее отклонение_____0.813486609581355
СКО/Среднее отклонение_____1.22927653414557
Медиана_____ -0.0612325915733022
Процентиль (0.00)_____ -2.42292817892483

Процентиль (0.25)	-0.704240233195885
Процентиль (0.50)	-0.0612325915733022
Процентиль (0.75)	0.648769025424276
Процентиль (1.00)	2.89464867248941
Ассиметрия	0.232263778185676
3*SQR(Da)	0.424264068711929
Эксцесс	-0.201306409267395
5*SQR(De)	1.41421356237309

ОПИСАТЕЛЬНАЯ СТАТИСТИКА для преобразованных данных:

Тип преобразования	натуральный логарифм (LN(X))
Число элементов	300
Сумма	1381.55212838541
Макс. значение	4.63212767668265
Мин. значение	4.58212305969478
Среднее значение	4.60517376128472
СКО среднего	0.00054347137079007
Границы среднего (0.95)	4.60427705212734, 4.6060704704421
Дисперсия	8.86083392605315E-05
Границы дисперсии (0.95)	7.59584965225791E-05, 0.000104722072221569
СКО	0.00941320026667506
Среднее отклонение	0.007661427478701
СКО/Среднее отклонение	1.22864835474121
Медиана	4.60464081075173
Процентиль (0.00)	4.58212305969478
Процентиль (0.25)	4.59856006863032
Процентиль (0.50)	4.60464081075173
Процентиль (0.75)	4.61131241531789
Процентиль (1.00)	4.63212767668265
Ассиметрия	0.207566482647164
3*SQR(Da)	0.424264068711929
Эксцесс	-0.21564570288542
5*SQR(De)	1.41421356237309



Корреляция

Данная процедура позволяет оценить силу линейной корреляционной связи между wybranнми факторами представленными полями таблицы MapInfo.

После выбора этой функции будет открыта форма выбора таблицы и полей таблицы (Рис. 8).

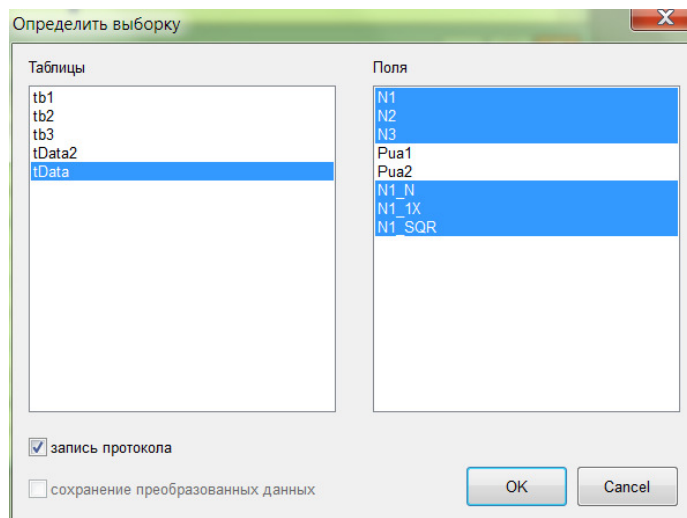
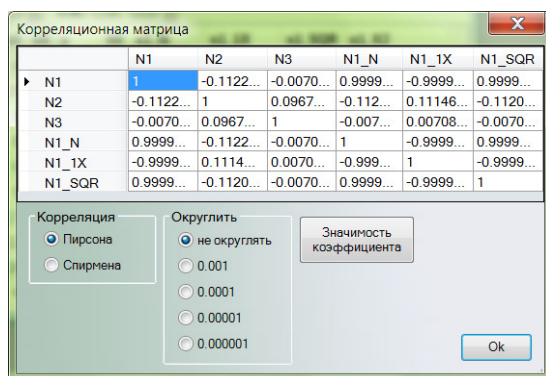
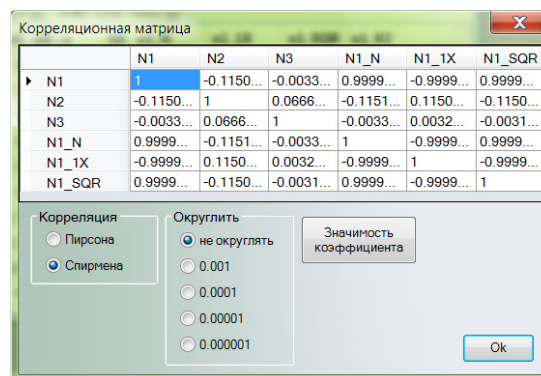


Рис. 8

Нажав кнопку «OK», перейдем к следующей форме представляющей результаты расчета (Рис. 9а). Выведенная корреляционная матрица представляет собой матрицу коэффициентов корреляции Пирсона для выбранных факторов. Выбрав «Корреляция Спирмена» получим матрицу коэффициентов корреляции Спирмена (Рис. 9б).



а



б

Рис. 9

Коэффициенты в таблице можно округлить до необходимого вида (Рис. 10а) или изменить размеры формы, чтобы увидеть значения в таблице с максимальным числом знаков. Также с этой целью можно изменить ширину конкретного столбца таблицы. При клике на заголовке столбца значения в таблице будут отсортированы по этому столбцу. Кнопка «Значимость коэффициента» позволяет оценить значимость выбранного коэффициента (Рис. 10б). В протокол выводятся полученные значения с краткими пояснениями.

Функции контекстного меню таблицы позволяют скопировать в буфер обмена выделенные значения или всю матрицу. Далее копию можно, например, вставить в документ Excel, Word и т.п.

Пример вывода в файл протокола приводится ниже. Имя файла протокола для этой функции имеет код *cm*, например *140806_210306_cm.html*.

Корреляционная матрица

	N1	N2	N3	N1_N	N1_1X	N1_SQR
N1	1	-0.1151	-0.0034	1	-1	1
N2	-0.1151	1	0.0667	-0.1152	0.115	-0.115
N3	-0.0034	0.0667	1	-0.0034	0.0033	-0.0032
N1_N	1	-0.1152	-0.0034	1	-1	1
N1_1X	-1	0.115	0.0033	-1	1	-1
N1_SQR	1	-0.115	-0.0032	1	-1	1

Корреляция
☐ Пирсона
☒ Спирмена

Округлить
☐ не округлять
☐ 0.001
☒ 0.0001
☐ 0.00001
☐ 0.000001

Значимость коэффициента

Ok

а

Корреляционная матрица

	N1	N2	N3	N1_N	N1_1X	N1_SQR
N1	1	-0.115	-0.003	1	-1	1
N2	-0.115	1	0.067	-0.115	0.115	-0.115
N3	-0.003	0.067	1	-0.003	0.003	-0.003
N1_N	1	-0.115	-0.003	1	-1	1
N1_1X	-1	0.115	0.003	-1	1	-1
N1_SQR	1	-0.115	-0.003	1	-1	1

Корреляция
☐ Пирсона
☒ Спирмена

Округлить
☐ не округлять
☒ 0.001
☐ 0.0001
☐ 0.00001
☐ 0.000001

Значимость коэффициента
0.953262509044552
0.476631254522276
0.523368745477724

Ok

б

Рис. 10

Коэффициент корреляции Пирсона (коэффициент корреляционного отношения Пирсона, выборочный коэффициент корреляции, коэффициент корреляции). Измеряет силу линейной корреляционной связи количественных признаков. Использование коэффициента корреляции в качестве меры связи оправдано лишь тогда, когда совместное распределение пары признаков стремится к нормальному распределению.

Коэффициент корреляции Спирмена (коэффициент корреляции рангов, показатель ранговой корреляции Спирмена). Может применяться при изучении линейной связи между рядами, представленными в количественной или порядковой шкале. Наиболее часто применяется при анализе корреляции порядковых признаков.

Корреляционная матрица

15 августа 2014 г., 23:01:04

Выборки из таблицы tData

поля: N1, N2, N3, N1_N, N1_1X, N1_SQR

Корреляционная матрица (корреляция Пирсона)

	N1	N2	N3	N1_N	N1_1X	N1_SQR
N1	1.0000	-0.1122	-0.0071	1.0000	-0.9999	1.0000
N2	-0.1122	1.0000	0.0967	-0.1122	0.1115	-0.1120
N3	-0.0071	0.0967	1.0000	-0.0071	0.0071	-0.0071
N1_N	1.0000	-0.1122	-0.0071	1.0000	-0.9999	1.0000
N1_1X	-0.9999	0.1115	0.0071	-0.9999	1.0000	-1.0000
N1_SQR	1.0000	-0.1120	-0.0071	1.0000	-1.0000	1.0000

Корреляционная матрица (корреляция Спирмена)

	N1	N2	N3	N1_N	N1_1X	N1_SQR
N1	1.0000	-0.1151	-0.0034	1.0000	-1.0000	1.0000
N2	-0.1151	1.0000	0.0667	-0.1152	0.1150	-0.1150
N3	-0.0034	0.0667	1.0000	-0.0034	0.0033	-0.0032
N1_N	1.0000	-0.1152	-0.0034	1.0000	-1.0000	1.0000
N1_1X	-1.0000	0.1150	0.0033	-1.0000	1.0000	-1.0000
N1_SQR	1.0000	-0.1150	-0.0032	1.0000	-1.0000	1.0000

Проверка значимости коэффициента корреляции (Спирмен):

$C(4,1) = 0.999999333325259$

1. Значение вероятности для двустороннего критерия (нулевая гипотеза: факторы имеют нулевую корреляцию)

0

2. Значение вероятности для левостороннего критерия (нулевая гипотеза: коэффициент корреляции больше или равен 0)

1

3. Значение вероятности для правостороннего критерия (нулевая гипотеза: коэффициент корреляции меньше или равен 0)

0

* Если полученное значение меньше заданного уровня значимости нулевая гипотеза отвергается.

Проверка значимости коэффициента корреляции (Спирмен):

$C(3,1) = -0.00339804304187887$

1. Значение вероятности для двустороннего критерия (нулевая гипотеза: факторы имеют нулевую корреляцию)

0.953262509044552

2. Значение вероятности для левостороннего критерия (нулевая гипотеза: коэффициент корреляции больше или равен 0)

0.476631254522276

3. Значение вероятности для правостороннего критерия (нулевая гипотеза: коэффициент корреляции меньше или равен 0)

0.523368745477724

* Если полученное значение меньше заданного уровня значимости нулевая гипотеза отвергается.

Аппроксимация полиномом

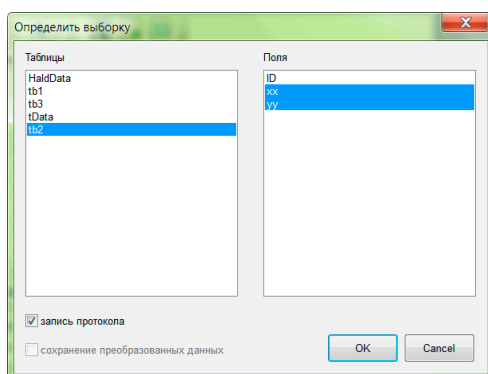
Понятие термина аппроксимация в математике трактуется достаточно широко, как замена одних математических объектов другими близкими, в каком-то смысле, к исходным объектам. Здесь и далее рассматривается аппроксимация зависимостей, приближение одних функций другими независимо от того известно аналитическое выражение для аппроксимируемой функции или нет⁶. Не рассматриваются распределения и статистические характеристики оценок, не вычисляются их доверительные интервалы и другие параметры стандартные для традиционного регрессионного анализа.

Принимается модель зависимости отклика y от значений фактора x в виде полинома

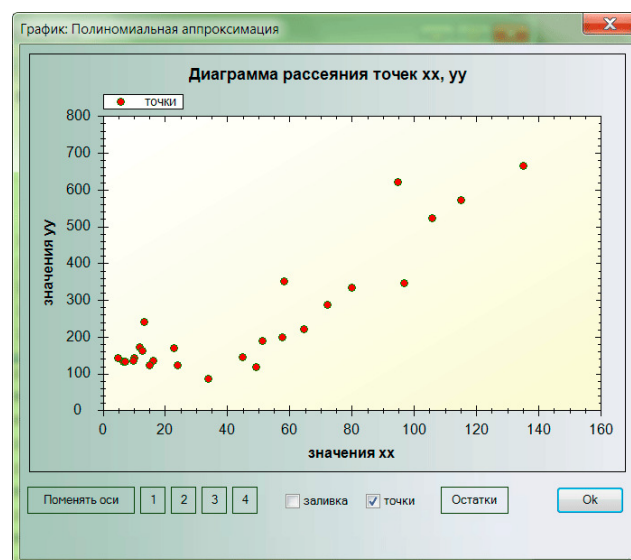
$$y(x) = \sum_{i=0}^t a_i x^i$$

Максимальное значение для t равно 4.

После выбора данной процедуры будет открыта форма определения выборок для фактора и отклика (Рис. 11а). На следующем этапе будет открыта новая форма (Рис. 11б) представляющая собой график рассеяния точек определяемых координатами {фактор, отклик}. По оси абсцисс откладываются значения фактора, по оси ординат значения отклика.



а



б

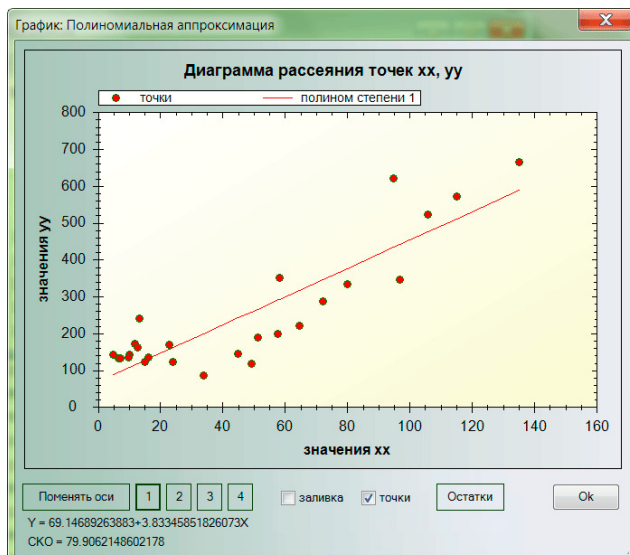
Рис. 11

Выбирая кнопки определяющие степень полинома (1, 2, 3, 4) получаем соответствующий график, найденное выражение для полинома и СКО (Рис. 12).

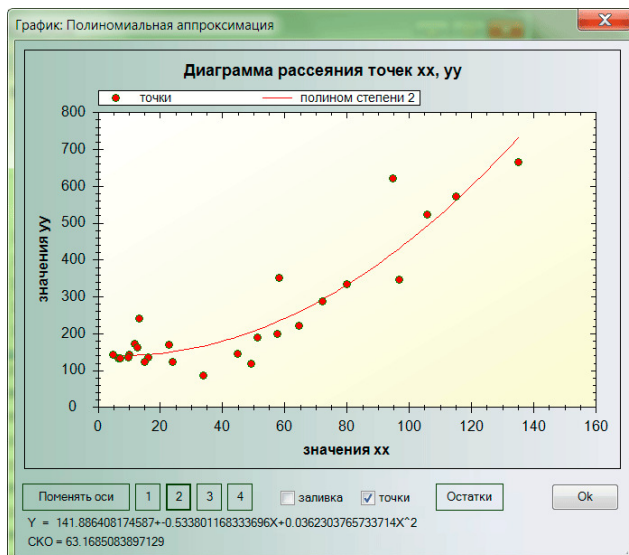
Если выбрать переключатель «заливка» то график примет вид как показано на Рис. 13а, если при этом отключить переключатель «точки» то график примет вид, показанный на Рис. 13б.

Кнопка «Остатки» открывает новую форму (Рис. 14) позволяющую проследить характер поведения остатков (e) определяемых как $e = y_{\text{выч.}} - y_{\text{табл.}}$. Информацию по остаткам можно сохранить в виде текстового файла (кнопка «Сохранить») представляющего собой набор записей следующего вида: x (фактор); $y_{\text{выч.}}$ (вычисленное значение отклика); e (остаток). Файл с указанным именем будет сохранен в папку *mi_Data*.

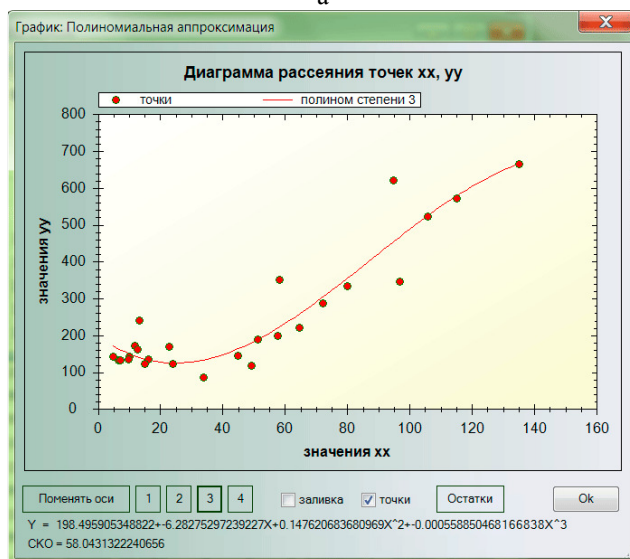
⁶ В первом случае целью обработки может быть упрощение известной функции.



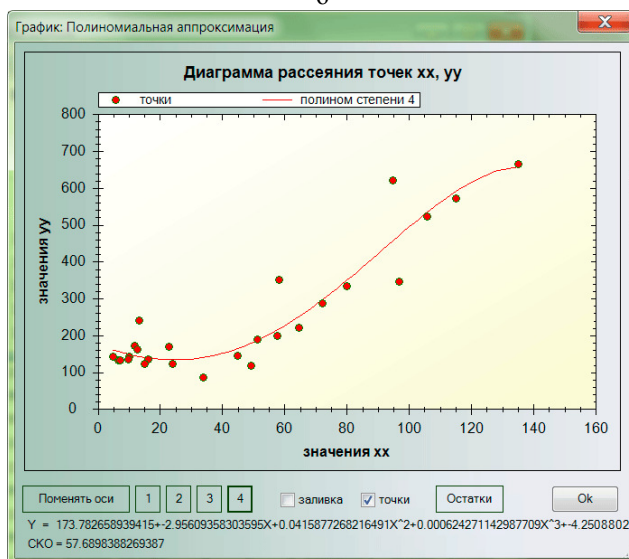
а



б

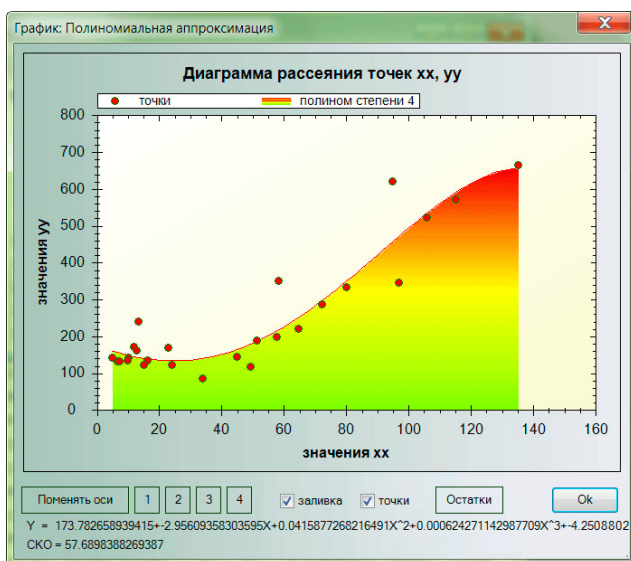


в

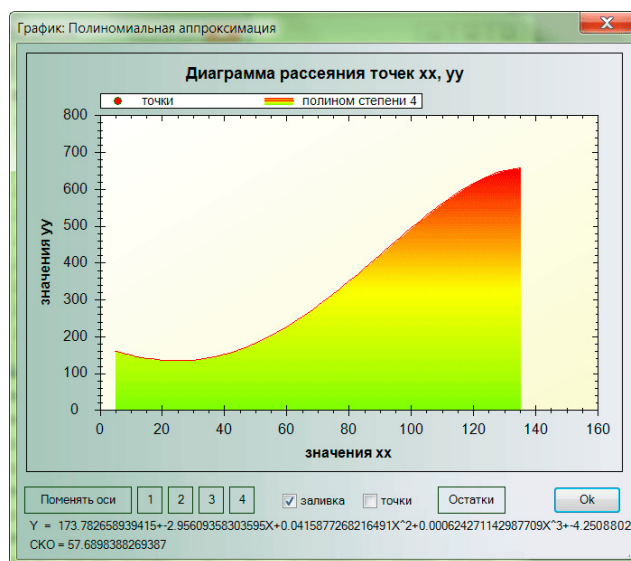


г

Рис. 12

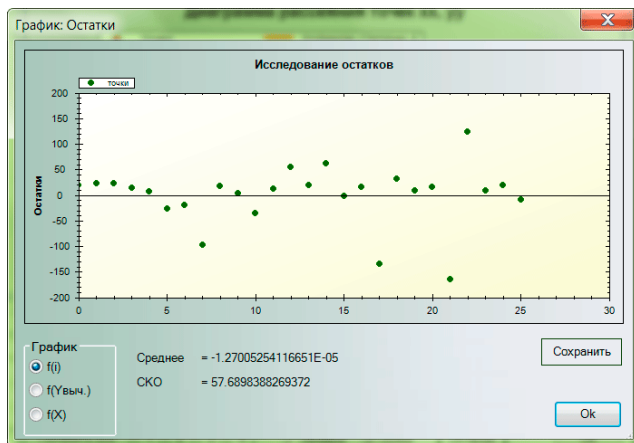


а

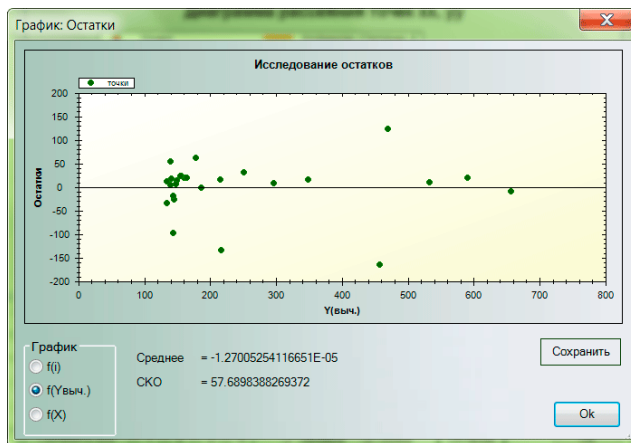


б

Рис. 13



а



б

Рис. 14

Кнопка «Поменять оси» позволяет поменять местами фактор и отклик. Приводимые ниже рисунки (Рис. 15, Рис. 16) иллюстрируют эту ситуацию, в качестве отклика используются значения из поля xx , а в качестве фактора значения из поля yy .

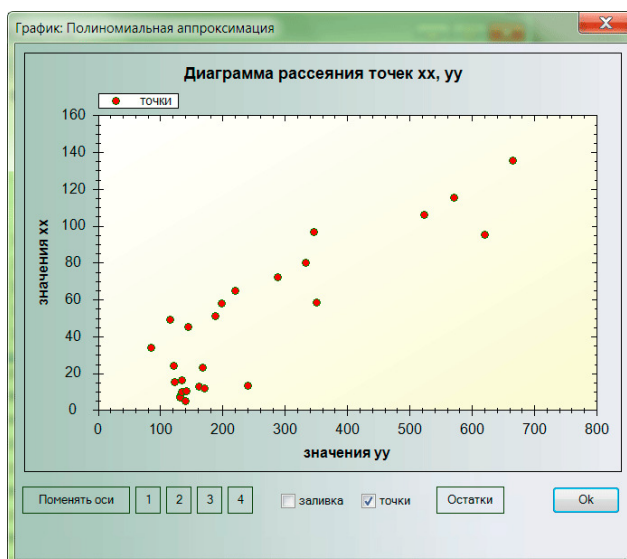
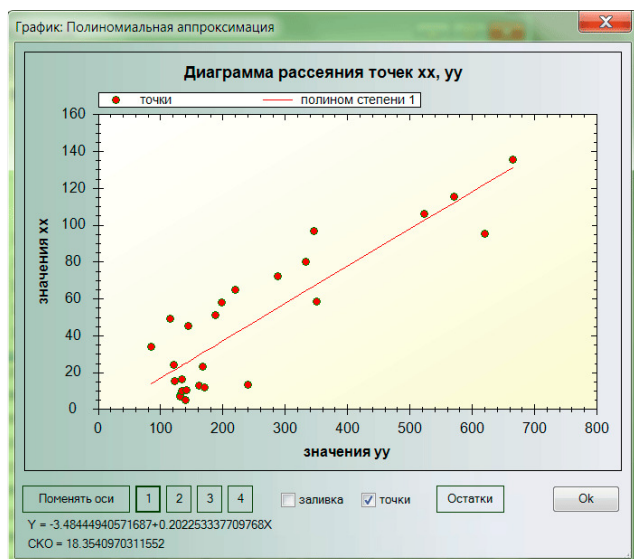
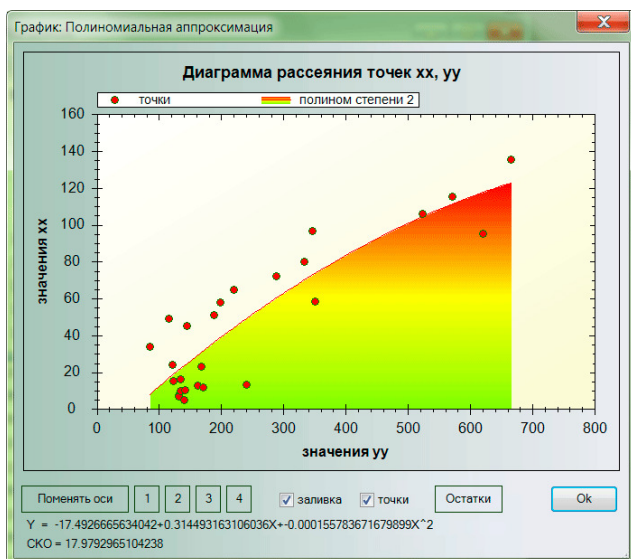


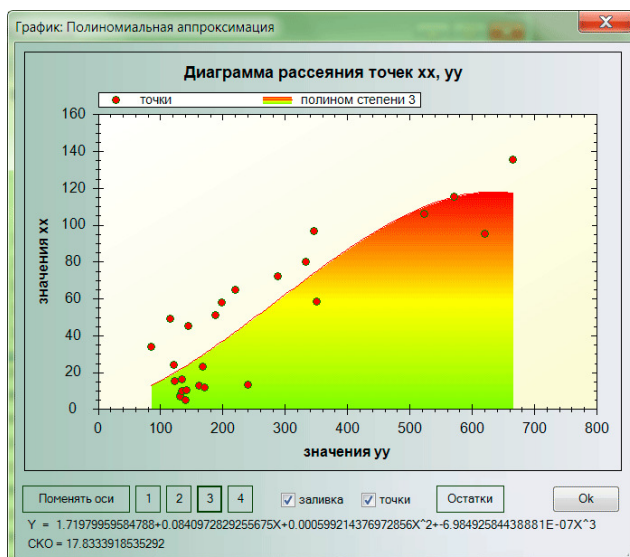
Рис. 15



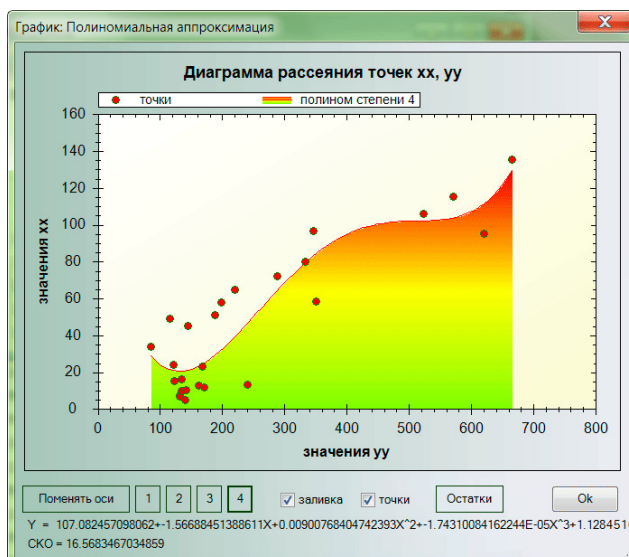
а



б



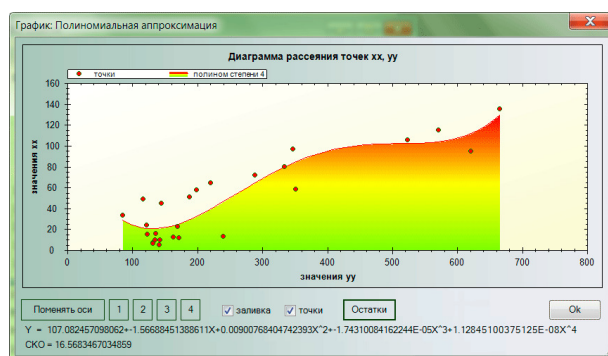
В



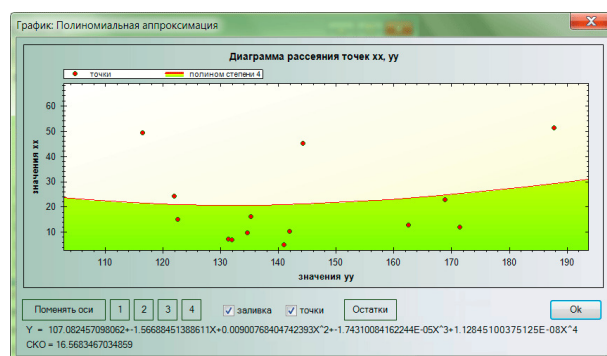
Г

Рис. 16

Размеры формы с диаграммой, как и формы с графиком остатков, могут быть изменены в обычном порядке, при этом изображение в поле графика масштабируется автоматически (Рис. 17а). Отдельную область графика при необходимости можно увеличить (Рис. 17б). После закрытия формы в окно «Сообщение» MapInfo выводится информация по выбранным полям (фактор, отклик) аналогичная показанной на Рис. 6а. Такая же информация присутствует и в файле протокола. Имя файла протокола для этой функции имеет код *apr*, например *140806_200506_apr.html*.



а



б

Рис. 17

Аппроксимация полиномом

17 августа 2014 г., 21:57:22

Выборки из таблицы tb2 поля: xx, yy

ОПИСАТЕЛЬНАЯ СТАТИСТИКА:

```

Таблица_____tb2
Поле_____xx
Число элементов_____26
Сумма_____1215.2
Макс. значение_____135.2
Мин. значение_____5
Среднее значение_____46.7384615384615
СКО среднего_____7.59405248441797
Границы среднего (0.95)_____33.766750946741, 59.7101721301821
Дисперсия_____1499.41046153846
Границы дисперсии (0.95)_____922.226760399299, 2857.16931971186
СКО_____38.722221805295
Среднее отклонение_____32.3491124260355
СКО/Среднее отклонение_____1.1970103320093
Медиана_____39.4
Процентиль (0.00)_____5
Процентиль (0.25)_____12.9
Процентиль (0.50)_____39.4
Процентиль (0.75)_____70.35
Процентиль (1.00)_____135.2
Ассиметрия_____0.672056447716548
3*SQR(Da)_____1.44115338424578
Эксцесс_____ -0.825458583652283
5*SQR(De)_____4.80384461415261

```

ОПИСАТЕЛЬНАЯ СТАТИСТИКА:

```

Таблица_____tb2
Поле_____yy
Число элементов_____26
Сумма_____6456.238
Макс. значение_____665.127
Мин. значение_____84.863
Среднее значение_____248.316846153846
СКО среднего_____33.0613916037186
Границы среднего (0.95)_____191.843335531823, 304.790356775869
Дисперсия_____28419.4459841354
Границы дисперсии (0.95)_____17479.652353099, 54154.0633681894
СКО_____168.580680933894
Среднее отклонение_____131.812633136095
СКО/Среднее отклонение_____1.2789417593975
Медиана_____170.18
Процентиль (0.00)_____84.863
Процентиль (0.25)_____134.84375

```

Процентиль (0.50) _____ 170.18
 Процентиль (0.75) _____ 322.3265
 Процентиль (1.00) _____ 665.127
 Ассиметрия _____ 1.27678809759528
 3*SQR(Da) _____ 1.44115338424578
 Экссесс _____ 0.28798427568349
 5*SQR(De) _____ 4.80384461415261

Аппроксимация полиномом степени 1: фактор (X) значения xx, отклик (Y) значения yy

$$Y = 69.14689263883 + 3.83345851826073X$$

$$СКО = 79.9062148602178$$

Аппроксимация полиномом степени 2: фактор (X) значения xx, отклик (Y) значения yy

$$Y = 141.886408174587 - 0.533801168333696X + 0.0362303765733714X^2$$

$$СКО = 63.1685083897129$$

Аппроксимация полиномом степени 3: фактор (X) значения xx, отклик (Y) значения yy

$$Y = 198.495905348822 - 6.28275297239227X + 0.147620683680969X^2 - 0.000558850468166838X^3$$

$$СКО = 58.0431322240656$$

Аппроксимация полиномом степени 4: фактор (X) значения xx, отклик (Y) значения yy

$$Y = 173.782658939415 - 2.95609358303595X + 0.0415877268216491X^2 + 0.000624271142987709X^3 - 4.25088023803255E-06X^4$$

$$СКО = 57.6898388269387$$

Информация по остаткам (X (фактор), Y(отклик вычисленный), E(остатки)) записана в файл C:\miStat\mi_Data\0001.txt

Аппроксимация полиномом степени 1: фактор (X) значения yy, отклик (Y) значения xx

$$Y = -3.48444940571687 + 0.202253337709768X$$

$$СКО = 18.3540970311552$$

Аппроксимация полиномом степени 2: фактор (X) значения yy, отклик (Y) значения xx

$$Y = -17.4926665634042 + 0.314493163106036X - 0.000155783671679899X^2$$

$$СКО = 17.9792965104238$$

Аппроксимация полиномом степени 4: фактор (X) значения yy, отклик (Y) значения xx

$$Y = 107.082457098062 - 1.56688451388611X + 0.00900768404742393X^2 - 1.74310084162244E-05X^3 + 1.12845100375125E-08X^4$$

$$СКО = 16.5683467034859$$

Информация по остаткам (X (фактор), Y(отклик вычисленный), E(остатки)) записана в файл C:\miStat\mi_Data\остатки_f025.txt




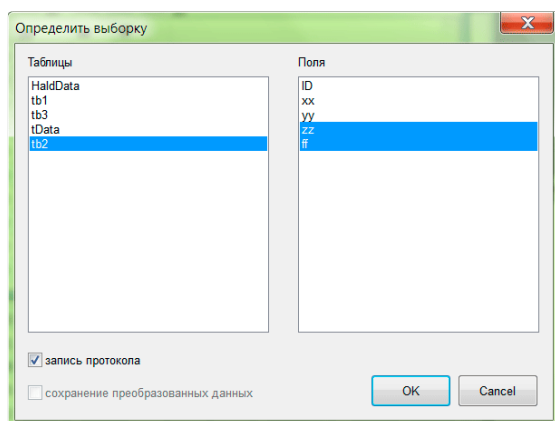
Нелинейная парная зависимость

В данной процедуре строятся различные формы нелинейной парной связи.

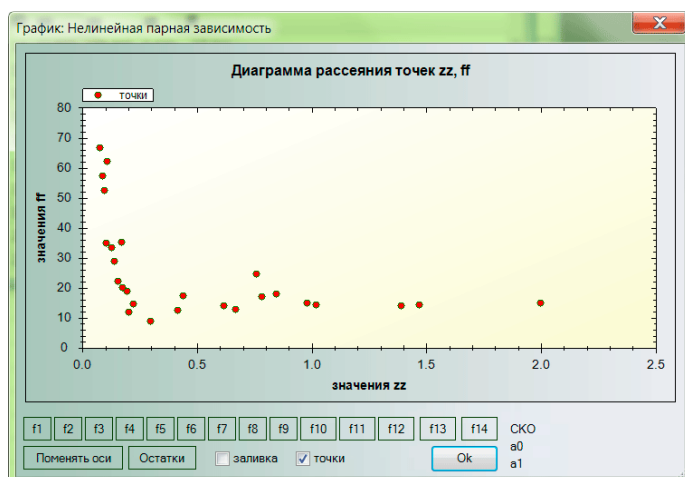
Таблица 1

Тип преобразования	Принятая модель	Ограничения
f1	$Y=a_0+a_1/X$	$X \neq 0$
f2	$Y=1/(a_0+a_1 \cdot X)$	$Y \neq 0$
f3	$Y=X/(a_0+a_1 \cdot X)$	$Y \neq 0$
f4	$Y=a_0 \cdot a_1^X$	$Y > 0$
f5	$Y=a_0 \cdot e^{(a_1 \cdot X)}$	$Y > 0$
f6	$Y=1/(a_0+a_1 \cdot e^{-X})$	$Y \neq 0$
f7	$Y=a_0 \cdot X^{a_1}$	$X > 0, Y > 0$
f8	$Y=a_0+a_1 \cdot \lg(X)$	$X > 0$
f9	$Y=a_0/(a_1+X)$	$Y \neq 0$
f10	$Y=a_0 \cdot X/(a_1 \cdot X+1)$	$X \neq 0, Y \neq 0$
f11	$Y=a_0 \cdot e^{(a_1/X)}$	$X \neq 0, Y > 0$
f12	$Y=a_0+a_1 \cdot X^2$	-
f13	$Y=a_0+a_1 \cdot X^3$	-
f14	$Y=a_0+a_1 \cdot X^4$	-

После выбора кнопки  будет открыта форма определения выборок для фактора и отклика (Рис. 18а). Далее, после нажатия кнопки «ОК», будет открыта новая форма (Рис. 18б) представляющая собой график рассеяния точек определяемых координатами {фактор, отклик}. По оси абсцисс откладываются значения фактора, по оси ординат значения отклика.



а



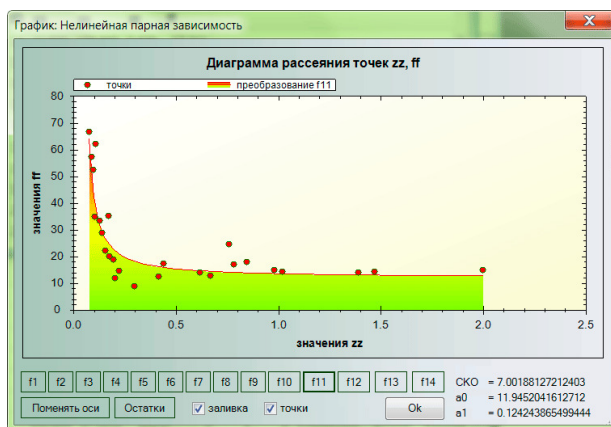
б

Рис. 18

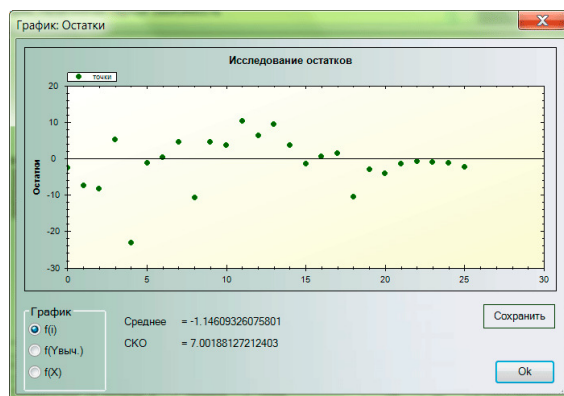
Выбор необходимого преобразования выполняется через кнопки $f1$, ..., $f14$, в соответствии с Таблица 1.

Далее на Рис. 19 показаны результат выбора преобразования f11 и соответствующий ему график остатков. Все результаты записываются в файл протокола⁷.

⁷ Если был выбран соответствующий переключатель на стадии выбора анализируемой таблицы.



а



б

Рис. 19

Кнопка «Поменять оси» позволяет поменять местами фактор и отклик. Приводимые ниже рисунки (Рис. 20, Рис. 21) иллюстрируют эту ситуацию, в качестве отклика теперь используются значения из поля zz, а в качестве фактора значения из поля ff. После закрытия формы в окно «Сообщение» MapInfo выводится информация по выбранным полям (фактор, отклик) аналогичная показанной на Рис. 6а. Такая же информация присутствует и в файле протокола. Имя файла протокола для этой функции имеет код *sr*, например *140806_221206_sr.html*.

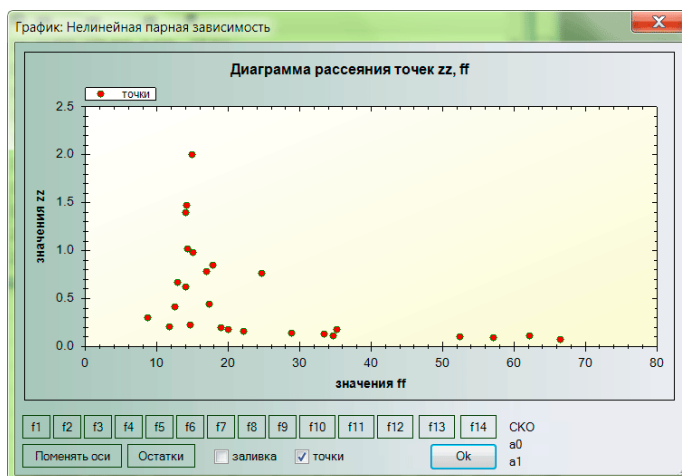
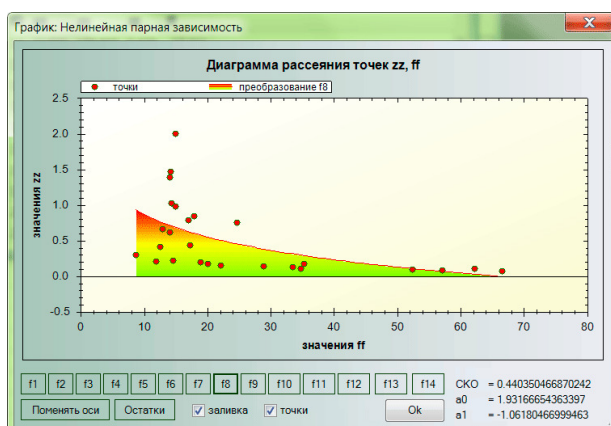
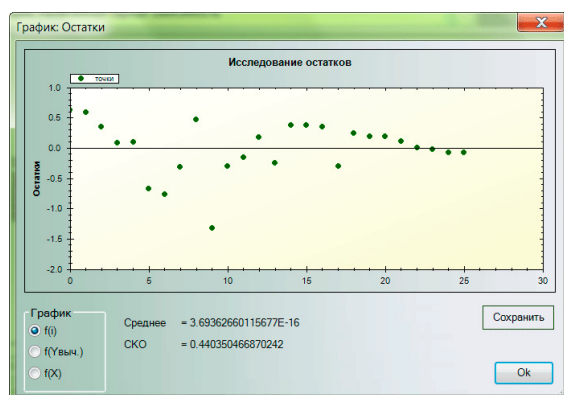


Рис. 20



а



б

Рис. 21

Если указанное пользователем преобразование не может быть выполнено, в силу установленных ограничений, будет выведено окно с соответствующим комментарием (Рис. 22).

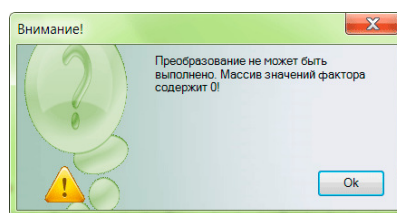


Рис. 22

Нелинейная парная зависимость

18 августа 2014 г., 21:48:38

Выборки из таблицы tb2 поля: zz, ff

ОПИСАТЕЛЬНАЯ СТАТИСТИКА:

Таблица_____tb2
Поле_____zz
Число элементов_____26
Сумма_____13.525
Макс. значение_____2
Мин. значение_____0.074
Среднее значение_____0.520192307692308
СКО среднего_____0.100208828403194
Границы среднего (0.95)_____0.349021523259515, 0.691363092125101
Дисперсия_____0.261087041538462
Границы дисперсии (0.95)_____0.160584084662982, 0.497508790283235
СКО_____0.510966771462158
Среднее отклонение_____0.409852071005917
СКО/Среднее отклонение_____1.24671023427568
Медиана_____0.259
Процентиль (0.00)_____0.074
Процентиль (0.25)_____0.14275
Процентиль (0.50)_____0.259
Процентиль (0.75)_____0.77525
Процентиль (1.00)_____2
Ассиметрия_____1.23993664368476
3*SQR(Da)_____1.44115338424578
Эксцесс_____0.724743430844625
5*SQR(De)_____4.80384461415261

ОПИСАТЕЛЬНАЯ СТАТИСТИКА:

Таблица_____tb2
Поле_____ff
Число элементов_____26
Сумма_____656.489
Макс. значение_____66.587
Мин. значение_____8.778
Среднее значение_____25.2495769230769
СКО среднего_____3.26844634564784
Границы среднего (0.95)_____19.666610494111, 30.8325433520428
Дисперсия_____277.751279373846
Границы дисперсии (0.95)_____170.833583694542, 529.262969876453
СКО_____16.6658716955893
Среднее отклонение_____12.9931834319527
СКО/Среднее отклонение_____1.28266269639547
Медиана_____17.605
Процентиль (0.00)_____8.778
Процентиль (0.25)_____14.22575

Процентиль (0.50)_____17.605
 Процентиль (0.75)_____32.3605
 Процентиль (1.00)_____66.587
 Ассиметрия_____1.28639804119043
 3*SQR(Da)_____1.44115338424578
 Эксцесс_____0.319276170571172
 5*SQR(De)_____4.80384461415261

Форма связи:

$$Y = a_0 + a_1/X$$

фактор (X) значения zz, отклик (Y) значения ff

параметры: $a_0 = 7.66280959194745$, $a_1 = 3.76385883104632$

СКО = 8.10620642497624

Форма связи:

$$Y = a_0 * X / (a_1 * X + 1)$$

фактор (X) значения zz, отклик (Y) значения ff

параметры: $a_0 = -205.425945300564$, $a_1 = -15.6446464109361$

СКО = 9.80902355341399

Форма связи:

$$Y = a_0 * e^{(a_1/X)}$$

фактор (X) значения zz, отклик (Y) значения ff

параметры: $a_0 = 11.9452041612712$, $a_1 = 0.124243865499444$

СКО = 7.00188127212403

Форма связи:

$$Y = a_0 * a_1^X$$

фактор (X) значения ff, отклик (Y) значения zz

параметры: $a_0 = 0.960157695707556$, $a_1 = 0.958044786184754$

СКО = 0.445130669504449

Форма связи:

$$Y = a_0 * e^{(a_1 * X)}$$

фактор (X) значения ff, отклик (Y) значения zz

параметры: $a_0 = 0.960157695707556$, $a_1 = -0.0428607524330755$

СКО = 0.445130669504449

Информация по остаткам (X (фактор); Y(отклик вычисленный); E(остатки))
 записана в файл C:\miStat\mi_Data\al.txt

Форма связи:

$$Y = a_0 + a_1/X$$

фактор (X) значения ff, отклик (Y) значения zz

параметры: $a_0 = 0.0298830484831494$, $a_1 = 9.17984260206359$

СКО = 0.457335055806984

Информация по остаткам (X (фактор); Y(отклик вычисленный); E(остатки))
 записана в файл C:\miStat\mi_Data\al3.txt



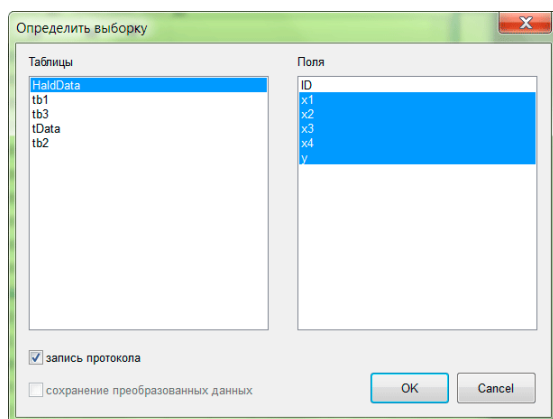
Многофакторная зависимость

Рассматривается модель вида

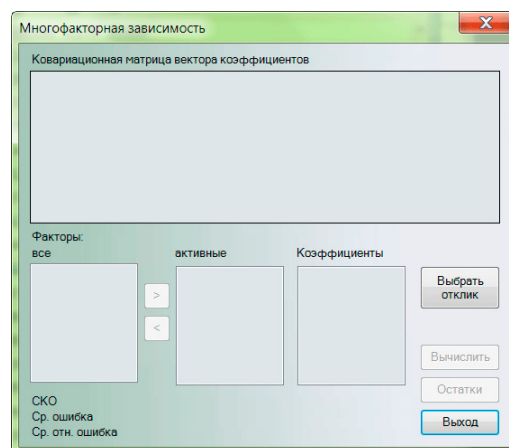
$$y(x) = \sum_{i=1}^n a_i x_i + a_0$$

где n – число факторов.

На первом шаге процедуры выбирается таблица и ее поля, определяющие факторы и отклик (Рис. 23а). Далее будет открыта основная форма для данной задачи (Рис. 23б).



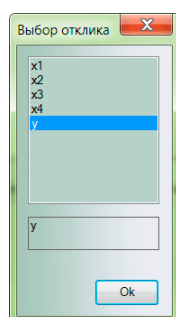
а



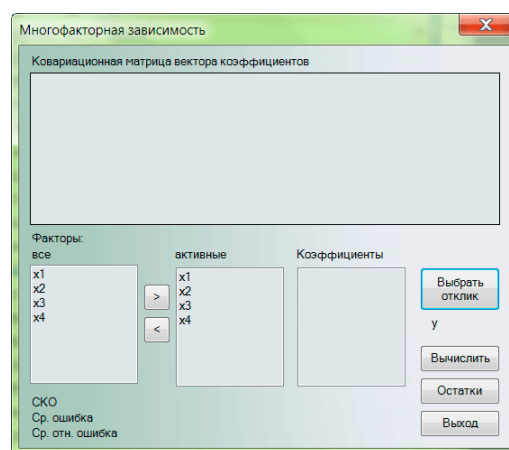
б

Рис. 23

Чтобы инициировать процесс нужно выбрать отклик (кнопка «*Выбрать отклик*»). После этого в открывшейся форме нужно выбрать поле представляющее отклик. На Рис. 24а это поле y . Теперь основная форма примет вид показанный на Рис. 24б. В списке «*активные*» показаны все используемые на данном этапе факторы. При необходимости этот список можно изменить: удалить факторы (кнопка $<$), добавить факторы (кнопка $>$).



а



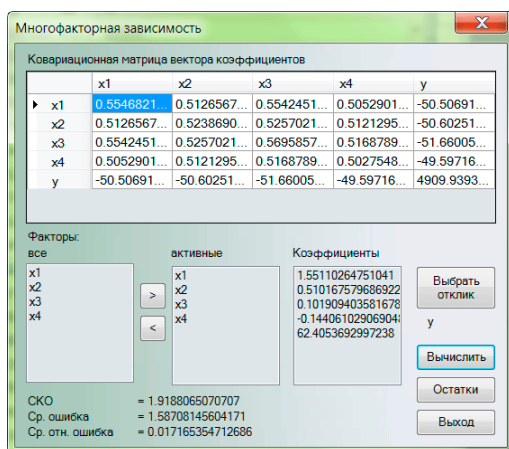
б

Рис. 24

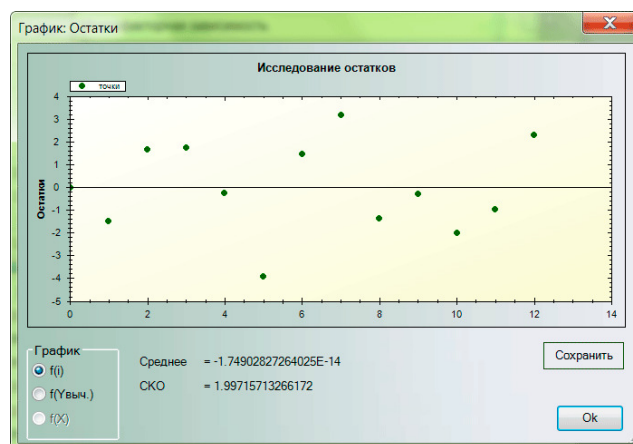
Процесс идентификации выбранной модели (подбора коэффициентов) запускается нажатием кнопки «*Вычислить*». Вид основной формы после проведения вычислений показан на Рис. 25а. Здесь в списке «*Коэффициенты*» размещаются вычисленные коэффициенты для выбранной модели. Коэффициенты соответствуют факторам в списке «*активные*». Последнее значение в списке это свободный член (a_0).

В верхней части формы размещена ковариационная матрица вектора коэффициентов. Последняя строка и последний столбец в ней, хотя они и обозначены именем отклика,

относятся к свободному члену. Напомним, что на диагонали этой матрицы располагаются дисперсии компонент вектора коэффициентов, а недиагональные элементы — ковариации между компонентами. Элементы матрицы или всю матрицу можно скопировать в буфер обмена через контекстное меню. При необходимости, выбрав кнопку «Остатки», можно проследить характер поведения остатков (Рис. 25б).



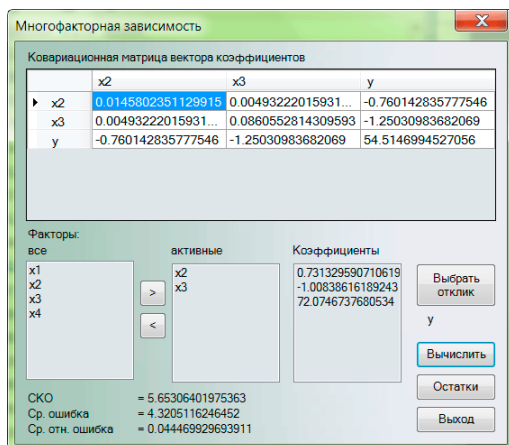
а



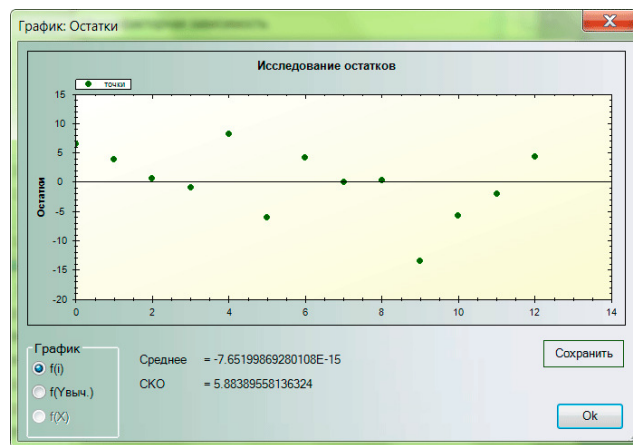
б

Рис. 25

Не выходя из основной формы можно изменить список активных факторов и выполнить вычисления для нового варианта (Рис. 26). При сравнении различных вариантов решения задачи можно ориентироваться, например на SKO.



а



б

Рис. 26

В том же сеансе работы, с выбранными ранее данными, можно перейти к решению новой задачи с новым откликом. Для этого нужно снова выбрать кнопку «Выборить отклик» и в открывшемся окне выбрать новый отклик (Рис. 27).

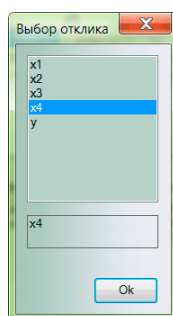
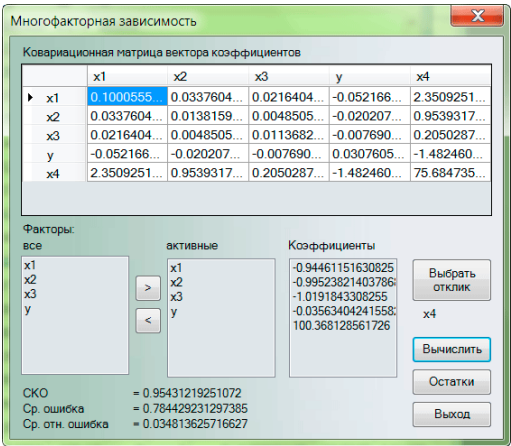
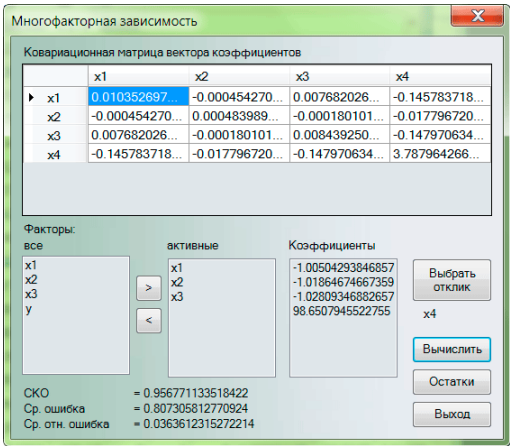


Рис. 27

Далее на Рис. 28 показаны два варианта решения для нового отклика (x4).
Имя файла протокола для этой функции имеет код *srr*, например *140806_181526_srr.html*.



а



б

Рис. 28

Многофакторная зависимость

19 августа 2014 г., 13:10:56

Выборки из таблицы HaldData

поля: x1, x2, x3, x4, y

Число наблюдений: 13

Определяется многофакторная зависимость вида $Y = a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_N \cdot X_N + a_0$

Корреляционная матрица (корреляция Пирсона)

	x1	x2	x3	x4	y
x1	1.000000	0.228579	-0.824134	-0.245445	0.730717
x2	0.228579	1.000000	-0.139242	-0.972955	0.816253
x3	-0.824134	-0.139242	1.000000	0.029537	-0.534671
x4	-0.245445	-0.972955	0.029537	1.000000	-0.821305
y	0.730717	0.816253	-0.534671	-0.821305	1.000000

Отклик: y

Факторы:

1. x1 _____ 1.55110264751041
2. x2 _____ 0.510167579686922
3. x3 _____ 0.101909403581678
4. x4 _____ -0.144061029069048
5. св. член _____ 62.4053692997238

СКО _____ = 1.9188065070707

Ср. ош. _____ = 1.58708145604171

Ср. отн. ош. ____ = 0.017165354712686

Ковариационная матрица вектора коэффициентов

	x1	x2	x3	x4	y
x1	0.554682	0.512657	0.554245	0.505290	-50.506920
x2	0.512657	0.523869	0.525702	0.512130	-50.602516
x3	0.554245	0.525702	0.569586	0.516879	-51.660056
x4	0.505290	0.512130	0.516879	0.502755	-49.597163
y	-50.506920	-50.602516	-51.660056	-49.597163	4909.939324

Отклик: y

Факторы:

1. x2 _____ 0.731329590710619
2. x3 _____ -1.00838616189243
3. св. член _____ 72.0746737680534

СКО _____ = 5.65306401975363

Ср. ош. _____ = 4.3205116246452

Ср. отн. ош. ____ = 0.044469929693911

Ковариационная матрица вектора коэффициентов

	x2	x3	y
x2	0.014580	0.004932	-0.760143
x3	0.004932	0.086055	-1.250310
y	-0.760143	-1.250310	54.514699

Отклик: x4

Факторы:

1. x1_____ -0.94461151630825
2. x2_____ -0.995238214037868
3. x3_____ -1.0191843308255
4. y_____ -0.0356340424155822
5. св. член_____ 100.368128561726

СКО_____ = 0.95431219251072

Ср. ош._____ = 0.784429231297385

Ср. отн. ош.____ = 0.034813625716627

Ковариационная матрица вектора коэффициентов

	x1	x2	x3	y	x4
x1	0.100056	0.033760	0.021640	-0.052166	2.350925
x2	0.033760	0.013816	0.004851	-0.020207	0.953932
x3	0.021640	0.004851	0.011368	-0.007691	0.205029
y	-0.052166	-0.020207	-0.007691	0.030761	-1.482460
x4	2.350925	0.953932	0.205029	-1.482460	75.684736

Отклик: x4

Факторы:

1. x1_____ -1.00504293846857
2. x2_____ -1.01864674667359
3. x3_____ -1.02809346882657
4. св. член_____ 98.6507945522755

СКО_____ = 0.956771133518422

Ср. ош._____ = 0.807305812770924

Ср. отн. ош.____ = 0.0363612315272214

Ковариационная матрица вектора коэффициентов

	x1	x2	x3	x4
x1	0.010353	-0.000454	0.007682	-0.145784
x2	-0.000454	0.000484	-0.000180	-0.017797
x3	0.007682	-0.000180	0.008439	-0.147971
x4	-0.145784	-0.017797	-0.147971	3.787964



Показать окно 'Сообщение'

Функции работы с окном «Сообщение» включены в данный набор инструментов т.к. это окно часто используется в программе для вывода информации.

Данная функция показывает (восстанавливает) окно «Сообщение» в правом нижнем углу экрана.



Очистить окно 'Сообщение'

Очищает окно «Сообщение» от текущего текста.



Выход

После выбора этой кнопки программа будет закрыта.