УДК 519.6

А.П. Гожий, И.И. Коваленко

СИСТЕМНОЕ ИСПОЛЬЗОВАНИЕ РОБАСТНЫХ И БУТСТРЕП МЕТОДОВ В ЗАДАЧАХ АНАЛИЗА ДАННЫХ

Введение

.

При применении статистических методов в задачах анализа данных в системах поддержки принятия решений и, в частности, при вычислении оценок параметров вероятностных распределений, проблема наличия в выборке аномальных измерений имеет важное значение. Присутствие единственного аномального наблюдения может приводить к оценкам, которые совершенно не согласуются с выборочными данными.

В борьбе с грубыми погрешностями измерений, если они не были обнаружены в процессе измерений, возможно использование трех подходов:

- исключение резко выделяющихся аномальных измерений из дальнейшей обработки;
- использование робастных методов обработки.
- использование бутстреп методов.

Большинство общепринятых статистических методов построено на предположении, что удовлетворяется ряд предпосылок, например, такие как независимость наблюдений, случайные величины имеют одинаковые средние и постоянные дисперсии, все наблюдения нормально распределены. Если эти предпосылки выполняются, то для оценки среднего используют среднеарифметическое. Эта оценка обладает свойствами несмещенности, состоятельности, достаточности. Даже при средних размерах объемов выборок для распределений нормального, пуассоновского и гамма распределений можно использовать асимптотически нормальные приближения. Однако, даже несколько резко выделяющихся значений, далеко отстоящих от основной массы наблюдений, могут изменить среднее значение очень сильно, а оценку дисперсии - катастрофически. Резко выделяющиеся наблюдения нарушают основное предположение метода наименьших квадратов - независимость дисперсии остатка от его математического ожидания. Действительно, резко выделяющееся наблюдение приводит к несимметричности распределения остатка, следовательно, условие независимости дисперсии и математического ожидания сразу нарушается. Обработка таких данных методом обычного регрессионного анализа может привести к ошибкам настолько большим, что полученная модель не будет иметь смысла[3,17].

Методы построения оценок статистик, которые малочувствительны к нарушениям исходных предпосылок, таких как вид распределения

 ISSN 1562-9945

и/или загрязнениями наблюдений посторонними помехами называют **робастными**[17,5]. Робастные методы устойчивы к ненормальности в хвостах нормальных распределений. К сожалению, нет единой робастной процедуры, оптимальной ко всем видам распределений и по отношению к конкретной статистике.

Для множественного регрессионного анализа робастность еще более желательная процедура, чем для оценки среднего. Ряд предположений регрессионного анализа таких, как нормальность, линейность, независимость, постоянство дисперсии могут рассматриваться лишь как средства более или менее удачного приближенного описания действительности. Следует отметить, что требование нормальности не входит в условия теоремы Гаусса-Маркова, объявляющей оценки МНК оптимальными среди всех линейных оценок. Дело здесь в том, что, как показал Фишер, всюду, за исключением очень малой окрестности нормального распределения остатков, все линейные оценки оказываются часто очень плохими [1]. Если в выборке имеются резко выделяющиеся наблюдения, то они несут с собой опасность при использовании классических статистических процедур. Даже одно единственное резко выделяющееся наблюдение может привести к непредсказуемым последствиям для оценки.

Главная особенность **бутстреп** методов состоит в том, что при их применении не нужна информация о виде закона распределения изучаемой случайной величины и в этом смысле методы могут рассматриваться как **непараметрические**, т.е они работают без опоры на существенную часть априорной информации, чем, по-видимому, и обусловлен такой выбор термина.

Бутстреп методы рассматриваются как методы управления выборкой в ходе обработки данных и отличается от традиционных методов тем, что он предполагает многократную обработку различных частей одних и тех же данных, как бы поворот их “разными гранями”, и сопоставление полученных таким образом результатов [5].

Постановка задачи. Целью данной работы является проведение анализа существующих методов робастного анализа и бутстреп методов и выбор наиболее эффективных и предпочтительных из них для процедур анализа данных, что облегчит их правильное применение в зависимости от решаемых прикладных задач.

Основные особенности применения робастных методов.

Для обеспечения надежности и эффективности оценок параметров регрессии нужно тем или иным способом нейтрализовать резко выделяющиеся наблюдения. Возможны следующие действия над этими наблюдениями: оставить все без изменений, ограничить влияние, жестко удалить, плавно удалить.

Любые действия ограничивающие влияние резко выделяющихся наблюдений предотвращают наихудший исход. Регрессионному анализу, построенному на методе наименьших квадратов, должен предшествовать тщательный анализ на содержание аномальных наблюдений.

Для вычисления параметров регрессионной модели без анализа резковыделяющихся данных, нужно пользоваться робастными методами, нечувствительными к нарушению основных предположений регрессионного анализа.

Выделим основные типы отклонений от строгих параметрических моделей. Выделяют три основных типа отклонений:

1. появление больших ошибок;
2. округление и группировка;
3. модель с самого начала выбиралась как некоторое приближение, например, в ее основу закладывалась центральная предельная теорема.

Относительно 1) необходимо отметить, что большие ошибки или искажения - это редкие явления (как правило, от 1 до 10 % в общем объеме данных), из-за каких-то неправильных действий. Поэтому относительно их можно указать на следующее:

- даже одна очень большая незамеченная ошибка способна совершенно обесценить статистический анализ (как это бывает при использовании метода наименьших квадратов);
- несколько процентов больших ошибок - скорее правило, чем исключение;
- современные робастные методы позволяют справиться с резко выделяющимися наблюдениями довольно просто и делают это даже лучше, чем классические объективные и субъективные методы отбраковки резко выделяющихся наблюдений.

Что касается 2), то все данные регистрируются и обрабатываются с ограниченной точностью, что делает их по большей части дискретными; далее их округляют, группируют или подвергают еще более грубой процедуре - классифицируют. В ряде ситуаций такие ошибки играют очень заметную роль: при очень грубой классификации, где непрерывное распределение служило бы очень плохим приближением; при изучении локально определенных величин вроде оценивания плотности и др.

Обратимся к 3). Даже большим множествам данных очень высокого качества (в плане однородности), не содержащем больших ошибок, свойственны небольшие, но заметные отклонения от нормальной модели, что выражается в наличии более длинных или более коротких “хвостов”. Указанные хвосты являются причиной того, что получаемые оценки параметров положения и масштаба таких распределений характеризуются определённой величиной смещения.

Основная идея процедур робастного (устойчивого) оценивания заключается в том, что имеется возможность получения несмещённых (малосмещённых) и эффективных оценок параметров основного распределения в условиях присутствия в исследуемых выборках так называемых “засоряющих” или “резковыделяющихся” значений.

 ISSN 1562-9945

Типичным примером при применении робастных методов [2] является случай, когда при анализе данных эксперимента по оценке прецизионности, значения стандартных отклонений повторяемости и воспроизводимости рассчитывают таким образом, что на них не влияет наличие выбросов. Если всех участников эксперимента можно разделить на два класса: производящих данные хорошего и плохого качества, то робастные методы дадут значения стандартных отклонений повторяемости и воспроизводимости, которые действительны для класса с хорошим качеством данных, и не окажут воздействия на данные плохого качества (при условии, что класс данных плохого качества не слишком велик).

Использование робастных методов для анализа данных не влияет на планирование, организацию или выполнение эксперимента по оценке прецизионности. Решение об использовании робастных методов или методов выявления и удаления выбросов должно приниматься экспертом по статистике и представляться в совет экспертов. При использовании робастных методов в ходе обработки данных необходимо, как и в других случаях, проводить тесты на наличие выбросов, проверку совместимости (однородности), а также исследовать причины отдельных выбросов. Однако сами исходные данные не должны исключаться как результаты этих измерений и проверок.

Данные эксперимента по оценке прецизионности позволяют рассчитать статистики двух типов:

а) средние значения в элементах, по которым рассчитывают стандартное отклонение, определяющее оценку межлабораторного расхождения;

б) стандартные отклонения или расхождения в пределах элементов (в том числе расхождения в эксперименте с распределенными уровнями), которые объединяют, чтобы получить оценку внутрилабораторного расхождения (вариации).

Робастные методы, не подменяют эти средние значения в элементах, стандартные отклонения или расхождения (или вариации), различия, а обеспечивают альтернативные способы их сочетания для получения статистик, используемых для расчетов стандартных отклонений повторяемости и воспроизводимости.

Робастные алгоритмы оценивания данных.

В настоящее время разработаны три класса робастных оценок [3,17]:

1. Устойчивые оценки на основе метода максимального правдоподобия (M-ценки);
2. Устойчивые оценки на основе ранговых критериев (R-оценки);
3. Линейные комбинации порядковых статистик (L-оценки).

Рассмотрим ряд процедур построения L-оценок, которые привлекают своей вычислительной простотой. Предварительно укажем на то, что для получения таких оценок необходимо преобразовать исходную выборку данных вида $x_1, x_2, \ldots, x_i, x_n$ в вариационный (ранжированный) ряд вида $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(i)} \geq x_{(n)}$ или $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(i)} \leq x_{(n)}$.

В качестве L-оценок, нацеленных на борьбу с указанными выше “хвостами”, рассмотрим следующие алгоритмы:

Алгоритм A

Этот алгоритм дает робастные величины среднего и стандартного отклонений данных, к которым он применяется, а именно:

a) средним значениям в элементах для любой модели;

b) расхождениям в элементах для модели с распределенными уровнями.

Обозначим индексом p общее число данных, расположенных в порядке возрастания: $x_1, x_2, +, x_i, +, x_p$.

Обозначим робастные среднее и стандартное отклонения этих данных x^*, s^*.

Рассчитаем первоначальные значения для x^*, s^*. в виде:

$$x^* = \text{медиана от } x_i (i = 1, 2, \ldots, p) \tag{1}$$

$$s^* = 1.483 \cdot \text{медиана от} |x_i - x^*| (i = 1, 2, \ldots, p) \tag{2}$$

Обновим значения x^*, s^*, как показано ниже.

Рассчитаем

$$\varphi = 1.5 s^* \tag{3}$$

Для каждого значения x_i $(i = 1, 2, \ldots, p)$ рассчитывают:

$$x_i^* = \begin{cases} x^* - \varphi, \text{если} x_i < x^* - \varphi, \\ x^* + \varphi, \text{если} x_i > x^* + \varphi, \\ x_i \varphi \text{- в остальных случаях} \end{cases} \tag{4}$$

Рассчитывают новые значения x^*, s^* по формулам:

$$x^* = \sum_{i=1}^{p} x_i^* / p, \tag{5}$$

$$s^* = 1.134 \sqrt{\sum_{i=1}^{p} (x_i^* - x^*)^2 / (p-1)} \tag{6}$$

Робастные оценки x^* и s^* могут быть получены итеративным расчетом, то есть повторением расчетов с (3) по (6) по несколько раз, до тех пор, пока изменения в оценках x^* и s^* от одного расчета до следующего станут минимальными. Этот метод прост для реализации.

Алгоритм S

Этот алгоритм применяют для внутрилабораторного стандартного отклонения (или внутрилабораторных расхождений) в любой модели эксперимента. Он дает робастное среднеквадратичное значение для стандартных отклонений или расхождений, к которым применен.

Обозначим индексом p общее число данных, распложенных в порядке возрастания: $w_1, w_2, +, w_i, +, w_p$.

(Это могут быть расхождения или стандартные отклонения).

 ISSN 1562-9945

Обозначим робастные среднеквадратичные значения w^*, а число степеней свободы, связанных с каждым w_i через ν. (Когда w_i – расхождение, $\nu = 1$. Когда w_i – стандартное отклонение из n результатов, $\nu = n - 1$). В таблице 1 находим соответствующие значения ξ и η, необходимые для использования алгоритма.

Таблица 1

Факторы, необходимые для робастного анализа. Алгоритм S

Степень свободы ν	**Ограничительный фактор** η	**Согласующий фактор** ξ
1	1.645	1.097
2	1.517	1.054
3	1.444	1.039
4	1.395	1.032
5	1.359	1.027
6	1.332	1.024
7	1.310	1.021
8	1.292	1.019
9	1.277	1.018
10	1.264	1.017

Найдем первоначальное значение для w^* в виде

$$w^* = \text{медиана от (середина по индексам) от} w_i (i = 1, 2, \ldots, p) \tag{7}$$

Обновляем величины w^* следующим образом.
Рассчитываем

$$\psi = \eta \times w^* \tag{8}$$

Для каждого w_i $(i = 1, 2, \ldots, p)$ рассчитываем

$$w_i^* = \begin{cases} \psi, \text{если} w_i > \psi \\ w_i, \text{в остальных случаях} \end{cases} \tag{9}$$

Рассчитываем новое значение w^* по формуле

$$w^* = \xi \sqrt{\sum_{i=1}^{p} (w_i^*)^2 / p} \tag{10}$$

Робастная оценка w^* может быть получена итеративным способом повторением расчетов с (8) по (10) по несколько раз, пока изменение оценки w^* от первого расчета до последующего станет минимальным. Это простой метод для реализации на компьютере.

Примечание: Для анализа данных описанными выше робастными методами применяют Алгоритм **A** к средним значениям элементов, а Алгоритм **S** – к каждой серии расхождений по очереди. Статистики, полученные в результате этих операций, используют затем для оценок стандартных отклонений повторяемости, промежуточной прецизионности и воспроизводимости.

Искажённые параметрические модели, помимо наличия “хвостов”, могут характеризоваться повышенными значениями величин асимметрии (A_S) и эксцесса (E_X).

Для оценивания параметра положения таких распределений можно воспользоваться адативными робастными процедурами, в основе которых лежит построение усечённых аналогов значений A_S^* и E_X^* по следующим формулам [4]:

$$A_S^* = \frac{[M(0.05) - L(0.05)]}{[M(0.5) - L(0.5)]}, \; E_X^* = \frac{[M(0.02) - L(0.02)]}{[M(0.5) - L(0.5)]}. \tag{11}$$

Здесь M(0.02), M(0.05), M(0.5) – среднее $\alpha \cdot n$ старших членов ранжированного ряда с константами усечения $\alpha_1 = 0.02$; $\alpha_2 = 0.05$; $\alpha_3 = 0.5$; L(0.02), L(0.05), L(0.5) – среднее $\alpha \cdot n$ младших членов ранжированного ряда с теми же значениями α.

В зависимости от полученных величин A_S* и E_X* можно реализовать две адаптивные процедуры построения и выбора усеченных средних с различными значениями α. Имеем:

$$X_{(A_S^*)} = \begin{cases} x^0(1/4), \text{ если } A_S^* < 2.0; \\ x(0), \text{ если } 2.0 \le A_S^* \le 4.0; \\ x(1/4) \text{ если } 2.0 < A_S^* \le 5.5; \\ x(1/2) \text{ если } A_S^* > 5.5. \end{cases} \tag{12}$$

Здесь $x^0(1/4)$ – среднее $\alpha \cdot n$ наибольших и $\alpha \cdot n$ наименьших членов ранжированного ряда; $x(0)$ – выборочное среднее по всему ряду; $x(1/4)$ – среднее внутренних членов ряда; $x(1/2)$ - выборочная медиана.

$$X_{(E_X^*)} = \begin{cases} x(1/4), \text{ если } E_X^* < 1.81; \\ x(1/6), \text{ если } 1.81 \le E_X^* \le 1.87; \\ x(3/8), \text{ если } E_X^* > 1.87. \end{cases} \tag{13}$$

Здесь $x(1/8), x(1/6), x(3/8)$ – усечённое среднее с константами усечения равными соответственно: $\alpha_1 = 0.125$; $\alpha_2 = 0.167$; $\alpha_3 = 0.375$.

Рассмотренные устойчивые оценки предназначены прежде всего для снижения величины смещения параметра положения (центра распределения).

Особенности применения бутстреп методов. Рассмотрим две бутстреп – процедуры, одна из которых основана на так называемом “размножении” выборок (resampling), а вторая получила в литературе название метод “складного ножа” [5, 6].

Обе процедуры нацелены всего на получение несмещенных оценок, которыми можно охарактеризовать устойчивость и достоверность какихлибо статистических выводов. При этом упор делается на случай, когда в распоряжении исследователя имеется единственная выборка и не сли-

 ISSN 1562-9945

шком большого объёма, что в ряде ситуаций характеризует процесс проведения различных этапов экспертизы.

Размножение выборки производится следующим образом. Берётся имеющаяся исходная выборка и исключается из неё один элемент, получается похожая выборка. Затем в её состав возвращается исключённый элемент и исключается следующий элемент, получается вторая похожая выборка. Поступив так со всеми элементами исходной выборки, можно получить столько выборок, похожих на исходную, каков её объём. Другими словами, если исходная выборке имеет объём n, то получается n “похожих” на неё выборок объёма $(n-1)$ каждая. Если же исключить по 2 наблюдения, то число “похожих” выборок возрастает до $n(n-1)/2$ объёма $(n-2)$ каждая. Далее необходимо обработать их тем же способом, что и исходную, и изучить устойчивость получаемых выводов – разброс оценок параметров, частота принятия или отклонения гипотез и т.д.

Необходимо отметить, что реализация бутстреп – процедур требует достаточно больших объёмов вычислений, однако рост вычислительных возможностей современных компьютеров наметил прогресс “бутстрепизации” статистических данных. Они начали широко применяться в задачах надёжности, прогнозирования, медицинской диагностики и др.

Рассмотрим один из возможных алгоритмов реализации описанной бутстреп – процедуры с целью изучения, разброса оценки выборочного среднего, характеризующего исходную выборку данных.

Имеется выборка вида $X_0 = x_1, x_2, x_3, x_i, \ldots, x_n$. Определяем по ней оценку выборочного среднего как $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$. Исключим из исходной выборки x_0 элемент x_1, тогда получим первую модифицированную выборку $X_1 = x_1, x_2, x_3, x_i, \ldots, x_n$ и $\bar{x}_1^* = \frac{1}{n-1}\sum\limits_{i=1}^{n-1} x_i$.

Затем исключим из X_o элемент x_2, а элемент x_1 вернём в её состав, получим $X_2 = x_1, x_2, x_3, x_i, \ldots, x_n$ и $\bar{x}_2^* = \frac{1}{n-1}\sum\limits_{i=1}^{n-1} x_i$ и т.д., пока не сформируем n модифицированных выборок с объёмом $(n-1)$ и n оценок выборочного среднего $\bar{x}_i^*,\ i = \overline{1,(n-1)}$.

Далее для опережения разброса (эффективности) полученных оценок $\bar{x}$ и x_i^* воспользуются методом наименьших квадратов и проанализируем остаточные разности по следующим формулам:

$$F_0 = \sum_{i=1}^{n}(x_i - \bar{x})^2,\ F_1 = \sum_{i=1}^{n}(x_i - x_1^*)^2,\ F_2 = \sum_{i=1}^{n}(x_i - x_2^*)^2\ \ldots$$

$$F_{n-1} = \sum_{i=1}^{n}(x_i - x_{n-1}^*)^2.$$

Кроме этого исследуем относительную эффективность рассмотренных оценок с использованием таких выражений:

$$G_1 = \frac{F_0}{F_1} = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{\sum\limits_{i=1}^{n} (x_i - x_1^*)^2},\ G_2 = \frac{F_0}{F_2} = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{\sum\limits_{i=1}^{n} (x_i - x_2^*)^2}, \ldots,$$

$$G_{n-1} = \frac{F_0}{F_{n-1}} = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{\sum\limits_{i=1}^{n} (x_i - x_{n-1}^*)^2}.$$

Последующий выбор показателей $F_i \to \min$ и $G_i \to \max$ позволит получить оценку выборочного среднего с минимальной величиной разброса, т.е. ту, которая характеризуется максимальной эффективностью.

Вторая бутстреп – процедура (типа “складывающаяся”) состоит в выборе некоторых опорных точек h_1 и h_2, которые представляют собой квант или кварт $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. После этого для оставшийся части выборки подсчитываются значения выборочного среднего и выборочной дисперсии по формуле:

$$\bar{X}_h = \frac{1}{n-2h} \sum_{i=h_1+1}^{h_2} x_{(i)};\ \sigma^2 = \frac{1}{n-2h-1} \sum_{i=h_1+1}^{h_2} (x_i - \bar{x})^2.$$

В дальнейшем описанная процедура повторяется до тех пор, пока из первоначального ряда не выделится устойчивая часть и минимальной величиной дисперсии.

Второй вид данной процедуры основан на преобразовании ряда, которые уменьшают его объём. Для этого от исходного ряда $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ переходят к новому со следующим набором данных:

$$\begin{cases} \frac{x_{(1)}+x_{(n)}}{2}, \frac{x_{(2)}+x_{(n-1)}}{2}, \ldots, \frac{x_{(n/2)}+x_{(n/2+1)}}{2}, \text{ если } n \text{ – четное;} \\ \frac{x_{(1)}+x_{(n)}}{2}, \frac{x_{(2)}+x_{(n-1)}}{2}, \ldots, x_{(n/2)}, \text{ если } n \text{ – нечетное.} \end{cases}$$

При этом вычисляются оценки выборного среднего и дисперсии $\bar{x}_{скл} = \frac{1}{m} \sum\limits_{j=1}^{m} x_j^{(1)}$ и $\sigma_{скл}^2 = \frac{1}{m} \sum\limits_{j=1}^{m} (x_j - \bar{x}_{скл})^2$, где m – количество членов модифицированной выборки.

Такая операция “складывания” выполняется несколько раз с проверкой на каждом шаге величины дисперсии $\sigma_{скл}^2$. По аналогии с первой процедурой устойчивая часть выборки также будет определяться минимальной величины $\sigma_{скл}^2$.

Результаты. Рассмотренный комплекс процедур статистического оценивания позволил сформировать системную технологию их выбора и использования, которая представлена на рис. 1.

В результате проведенного анализа робастных и бутстреп методов оценивания и технологии их выбора и использования был разработан экпериментальный программный комплекс робастного и бутстреп оценивания для систем поддержки принятия решений.

 ISSN 1562-9945

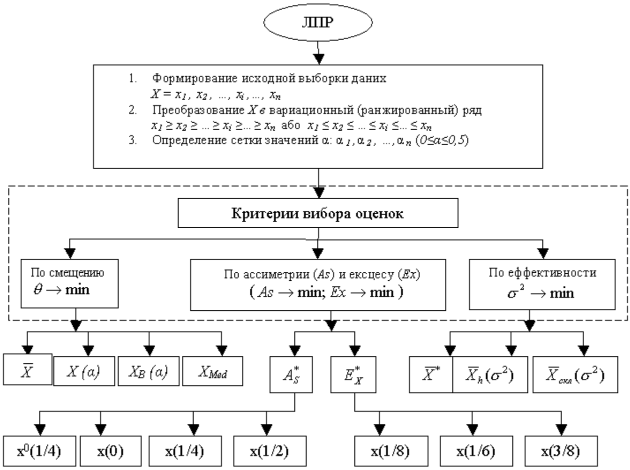

Рис. 1 – Структура системной технологии применения методов.

Исходные данные загружаются в виде текстового файла со специальным шаблоном для занесения данных затем выбирается метод оценивания (рис.2).

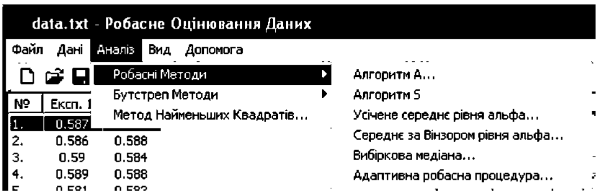

Рис. 2 – Выбор метода оценивания в программном комплексе

Такая структура файла достаточно удобна, поскольку в таком случае файл можно формировать не только в программе, но и с помощью других программных средств, которые выдают в качестве исходного результата статистические данные. Это делает созданную программу более гибкой в применении с другими программными комплексами. Программа позволяет добавлять экспериментальные данные, то есть каждый експерт может оценить данные несколько раз. Результаты оценивания выводятся на екран, либо в виде отдельного файла.

С помощью разработанного программного комплекса проводилось оценивание в задачах принятия решений при покупке недвижимости. Результаты оценивания эффективны.

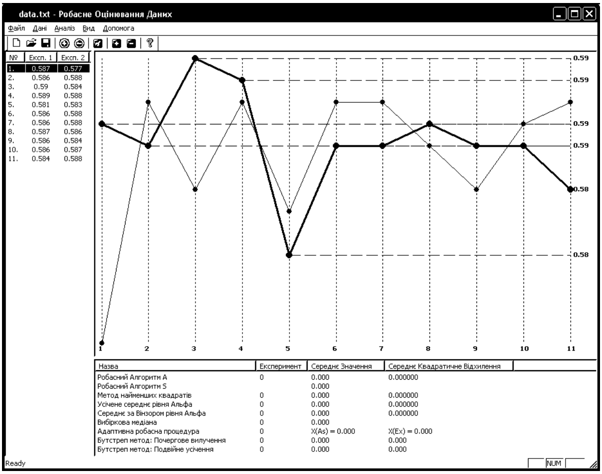

Рис. 3 – Результаты работи программного комплекса.

Литература

1. ГОСТ 11.011-83. Прикладная статистика. Правила определения оценок и доверительных границ для параметров гамма-распределения. - М.: Изд-во стандартов, 1984. - 53 с. - Переиздание: М.: Изд-во стандартов, 1985. - 50 с.
2. Рао С.Р. Линейные статистические методы и их применения. – М.: Наука, 1968. - 548 с.
3. Вентцель Е.С. Теория вероятностей. – М.: Наука, 1964. -576 с.
4. Коваленко И. И. Адаптивный алгоритм повышения надежности экспертных оценок // Адаптивні системи автоматичного управління. – Дніпропетровськ: „Системні технології”. – 2000. – вип. 3 (23). – с.132-136.
5. Орлов А.И.Устойчивость в социально-экономических моделях.– М.:Наука, 1979.– 396 с.
6. Ефрон Б. Нетрадиционные методы многомерного статистического анализа .– М.:Финансы и статистика, 1988.– 263 с.
7. Закс Ш. Теория статистических выводов. – М.: Мир, 1975. – 776 с.
8. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. – М.: Наука, 1979. – 528 с.
9. Боровков А.А. Математическая статистика / Учебное пособие для вузов. – М.: Наука, 1984. – 472 с.
10. Кендел М.Дж., Стьюарт А.Статистические выводы и связи.– М.: Наука, 1973. – 900 с.

 ISSN 1562-9945

11. Петрович М.Л., Давидович М.И. Статистическое оценивание и проверка гипотез на ЭВМ. – М.: Финансы и статистика, 1989. – 191 с.
12. Кендэл М. Ранговые корреляции. - М.: Статистика, 1975. - 216 с.
13. Большев Л.Н.,Смирнов Н.В.Таблицы математической статистики.– М.: Наука, 1983. - 416 с.
14. Себер Дж. Линейный регрессионный анализ. - М.: Мир, 1980.- 456 с.М.: Изд-во "Экзамен", 2003.–576с.
15. Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. - М.: Наука, 1976. – 736 с.
16. Демиденко Е.З. Линейная и нелинейная регрессии. - М.:Финансы и статистика, 1981. - 302 с.
17. Хьюбер П. Робастность в статистике: Пер. с англ. - М.: Мир, 1984. - 304 с.