

**Степаньян Владимир Владимирович**,  
студент экономико-технологического факультета НОУ ВПО «Международный инновационный университет», г. Сочи  
[step.wo@mail.ru](mailto:step.wo@mail.ru)



## **Алгоритм и программа обработки массивов ненормированных данных о состоянии образовательных систем методом планирования эксперимента**

**Аннотация.** При исследовании сложных многофакторных процессов, в которых на главный выходной параметр влияет множество факторов, применяется теория планирования эксперимента. При этом значения факторов задаются фиксированно в обе стороны от номинального значения с определенным интервалом варьирования. Предлагаются алгоритм и программа обработки массивов факторов с любыми отклонениями от номинала, полученными в ходе эксперимента. В основу положены оригинальные подходы, разработанные ранее. Указанная программа может быть использована при оптимизации физических, технических, экономических и социальных процессов.

**Ключевые слова:** алгоритм, вычисление, коэффициент, матрица, функция, уровень регрессии, качество, газ.

**Раздел:** (01) педагогика; история педагогики и образования; теория и методика обучения и воспитания (по предметным областям).

При исследовании физических, технических, экономических и социальных процессов существует проблема обработки массивов, размеры которых могут превышать десятки тысяч строк. Обработка таких массивов вручную, с использованием известных пакетов прикладных программ (ППП), например MathCAD, практически невозможна, так как из этого огромного количества строк необходимо реализовать полный факторный эксперимент. Даже массив, состоящий из тысячи строк, может обрабатываться вручную несколько дней. Для того чтобы решить данную проблему, мы разработали программу статистической обработки массивов непрерывных данных методом планирования эксперимента [1, 2].

Алгоритм работы программы содержит четыре основных этапа и представлен на рисунке. На первом этапе отсеиваются все значения факторов, не входящих в нормальный закон. Это позволяет исключить из статистического ряда грубые ошибки, за счет чего размер массива значительно сокращается. На втором этапе производится расчет коэффициентов изменения факторов и реализуются методы повышения адекватности уравнения регрессии, разработанные при участии автора и изложенные в [2, 3]. На третьем этапе реализуется полный факторный эксперимент (ПФЭ) типа  $2^n$ , где  $n$  – число факторов. На четвертом этапе получаем уравнение регрессии с проверкой адекватности по критерию Фишера и отсеиванием незначущих факторов по критерию Стьюдента.

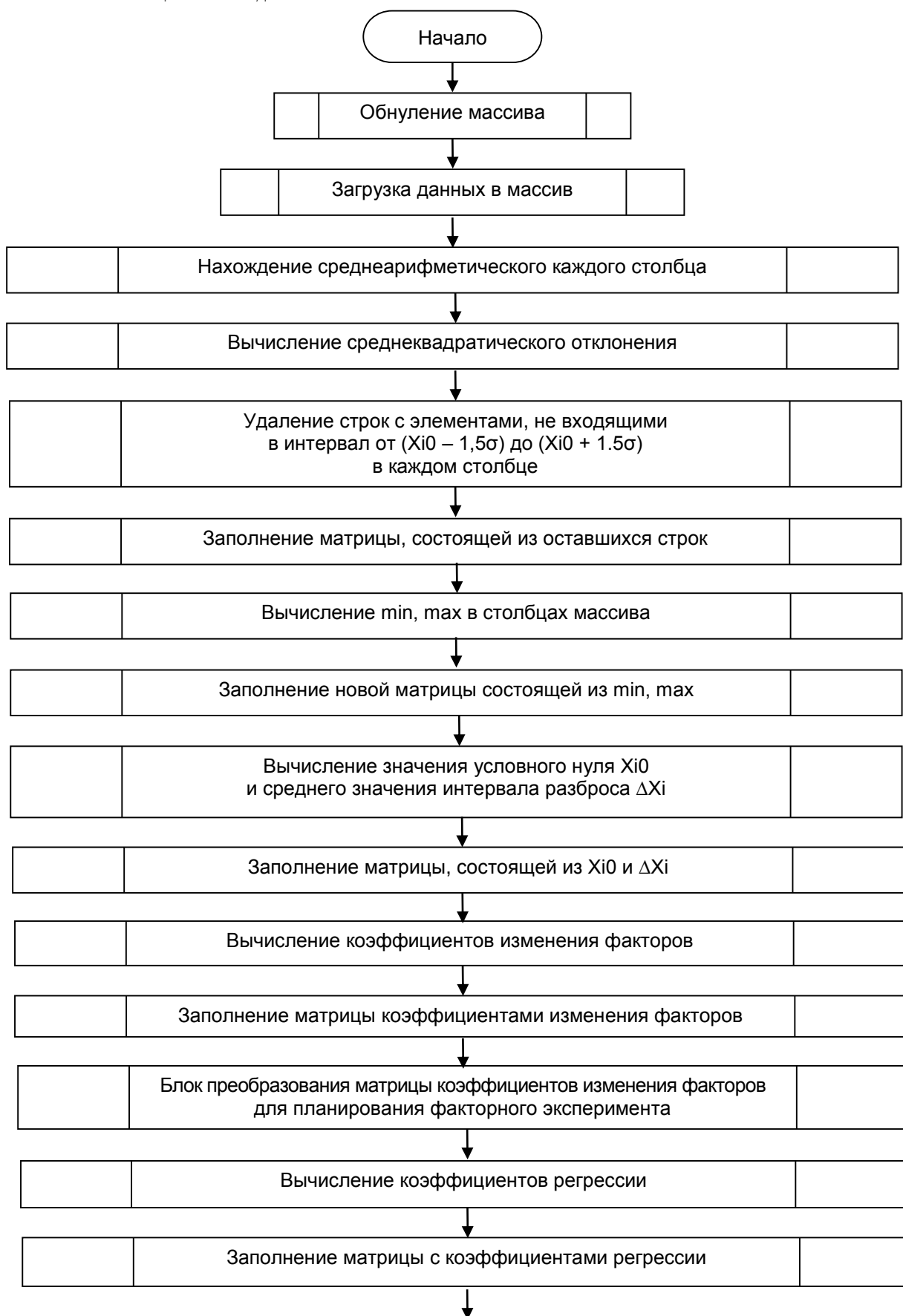
Рассмотрим предлагаемый алгоритм более подробно.

### **Обнуление массива**

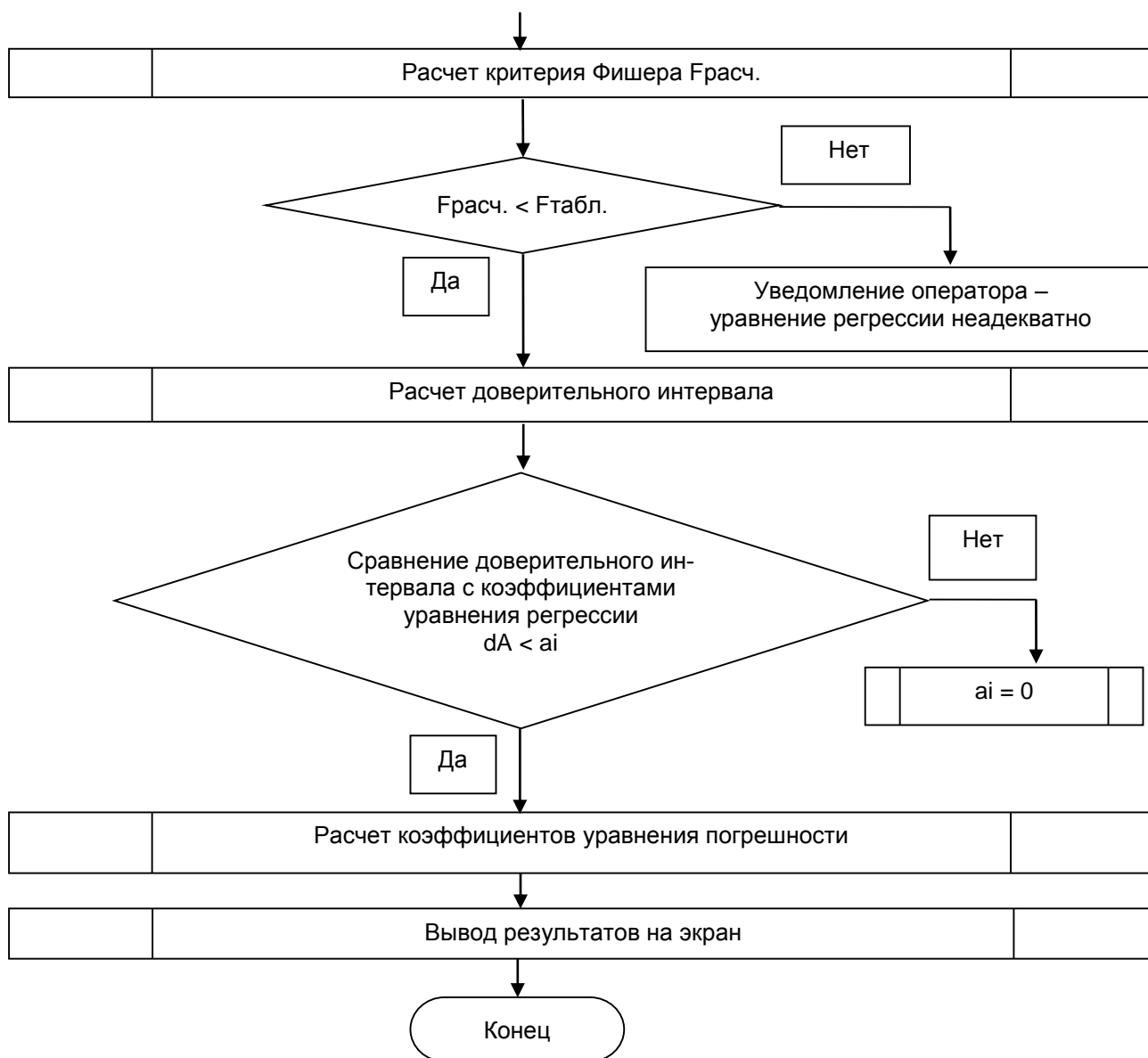
При загрузке программы происходит обнуление массива для возможности дальнейшего заполнения матрицы исходного массива данных.

#### **1) Загрузка данных в массив**

Загрузка данных в массив производится либо поэлементным введением в матрицу, либо загрузкой из файла с расширением \*.bak. Эту операцию осуществляет модуль сохранения и загрузки.



Алгоритм работы программы



Алгоритм работы программы (продолжение)

**2) Нахождение среднеарифметического каждого столбца**

1. Находим сумму всех элементов по каждому столбцу.
2. Делим каждое полученное значение на количество строк.
3. Вычисляем квадрат разности  $(Y_{ij} - \hat{Y})^2$ .
4. Присваиваем полученные значения новой матрице.

**3) Вычисление среднеквадратического отклонения  $\sigma$**

1. Вычисляем сумму по каждому столбцу.
2. Полученную сумму делим на количество строк.
3. Присваиваем полученные значения новой матрице.

**4) Удаление строк с элементами, не входящими в интервал от  $(X_{i0} - 1,5\sigma)$  до  $(X_{i0} + 1,5\sigma)$  в каждом столбце**

1. Определяем границу  $(X_{i0} - 1,5\sigma)$  по каждому столбцу.
2. Определяем границу  $(X_{i0} + 1,5\sigma)$  по каждому столбцу.
3. Циклически сравниваем каждое значение каждого столбца с полученным интервалом.

4. Удаляем полностью строку, если элемент какого-либо столбца не входит в этот интервал.

5. Количество строк присваиваем (Количество строк – 1) – «kolstr:=kolstr-1».

**5) Вычисление min, max в столбцах массива и заполнение новой матрицы, состоящей из min, max**

1. Декларируем некоторые значения MAX и MIN (бесконечно большое и бесконечно маленькое).

2. Перебираем все элементы матрицы M по столбцам.

3. Если значение элемента матрицы больше значения MAX, то присваиваем MAX значение этого элемента.

4. Если значение элемента матрицы меньше значения MIN, то присваиваем MIN значение этого элемента.

5. Заносим получившиеся значения MIN и MAX для каждого столбца в матрицу минимальных и максимальных значений.

6. Возвращаем функции значение матрицы минимальных и максимальных значений.

**6) Вычисление значения условного нуля  $X_{i0}$ , среднего значения интервала разброса  $\Delta X_i$  и заполнение матрицы, состоящей из  $X_{i0}$  и  $\Delta X_i$**

1. Перебираем все значения матрицы M по столбцам.

2. Получаем из матрицы M значения минимумов и максимумов.

3. Производим вычисление значений условного нуля и среднего значения интервала разброса по формулам и заносим данные значения в матрицу.

4. Возвращаем функции значение матрицы значений условного нуля и среднего значения интервала разброса.

Вычисление значений условного нуля и среднего значения интервала разброса вычисляется по следующим формулам соответственно [1]:

$$X_{i0} = \frac{X_{\max_i} + X_{\min_i}}{2}; \Delta X_i = \frac{X_{\max_i} - X_{\min_i}}{2} \quad (1)$$

**7) Вычисление коэффициентов изменения факторов и заполнение матрицы коэффициентами изменения факторов**

**Начало**

1. Перебираем все значения матрицы M по столбцам.

2. Производим вычисление коэффициентов изменения факторов по формуле, используя значения коэффициентов начальной матрицы, а также значений условного нуля и среднего значения интервала разброса. Заносим коэффициенты в матрицу.

3. Возвращаем функции значение получившейся матрицы.

Вычисление коэффициентов изменения факторов производится по формуле [2]:

$$K_{ij} = \frac{X_{ij} - X_{i0}}{\Delta X_i} \quad (2)$$

**8) Блок преобразования из матрицы коэффициентов изменения факторов в матрицу планирования полного факторного эксперимента**

Данный блок является наиболее важной частью программы, так как здесь идет преобразование полученной матрицы до матрицы полного факторного эксперимента, т. е. из тысячи строк необходимо выделить такие строки, чтобы окончательная матрица обладала следующими свойствами [1]:

$$\sum_{j=1}^m x_{ij} = 0; \sum_{j=1}^m x_{ij}^2 = m; \sum_{j=1}^m x_{iu} x_{ju} = 0. \quad (3)$$

Данный блок разделяется на 3 раздела, описания которых представлены ниже.

### Первый раздел:

1. Перебираем последовательно все строки матрицы М.
2. Проверяем каждую строку на условие соответствия виду: ( -, -, -, - ); (+, +, +, +); (+, -, -, -); (-, +, +, +); (-, -, +, -); (+, -, -, +); (-, +, -, +); (+, -, +, -); (-, -, -, +); (+, -, -, -); (-, -, +, +); (+, +, -, -); (-, -, +, +); (+, +, -, +); (-, +, -, -); (+, -, +, +).
3. Удаляем все строки вида ( -, -, -, - ) и (+, +, +, +).
4. Если значения элементов строки не удовлетворяют данному условию, то вызываем процедуру удаления данной строки.

### Второй раздел:

1. Перебираем последовательно все строки матрицы М.
2. Считаем количество строк каждого вида.
3. Считаем количество строк, не имеющих пары.
4. Если строка не имеет пары, то вызываем процедуру удаления строки.

### Третий раздел

Проверяем: если количество строк больше 16, тогда удаляем парные строки, значение коэффициентов изменения факторов по модулю которых больше. Таким образом, остается только 16 строк и полученная матрица удовлетворяет условиям (3).

### 9) Вычисление коэффициентов регрессии и заполнение матрицы с коэффициентами регрессии

1. Перебор матрицы К по строкам.
2. Вычисление коэффициентов по формулам.
3. Занесение коэффициентов в матрицу.
4. Присвоение функции значения полученной матрицы.

Вычисление коэффициентов регрессии выполняется по формулам [1]:

$$a_0 = \frac{1}{m} \sum_{j=1}^m y_j; a_i = \frac{1}{m} \sum_{j=1}^m K_{ji} y_j. \quad (4)$$

### 10) Расчет критерия Фишера Фрасч. и доверительного интервала

Полученное уравнение регрессии необходимо проверить на адекватность по критерию Фишера, а коэффициенты уравнения регрессии – на значимость по критерию Стьюдента. Для этого производим следующие операции.

1. Перебор матрицы по строкам.
2. Вычисление значений Фрасч. и dAi с использованием значений коэффициентов из матрицы.
3. Присваивает функции значение Фрасч. и dAi.
4. Проверка уравнения регрессии на адекватность.

В этой процедуре выполняется расчет следующих промежуточных величин, необходимых для вычисления значения Фрасч. и dAi: Двоспр. – дисперсия воспроизводимости; Дад – дисперсия адекватности; f – число степеней свободы  $\hat{Y}_i$  – значение выходного параметра, вычисленное по уравнению регрессии.

Сначала определяется число степеней свободы f как [1]:

$$f = m - (n + 1), \quad (5)$$

где m – число опытов в матрице факторного эксперимента; n – число факторов.

Затем вычисляется  $\hat{Y}_i$  – значение выходного параметра, вычисленное по уравнению регрессии [1]:

$$Y = a_0 + a_1 * q_1 + a_2 * q_2 + a_3 * q_3 + a_4 * q_4, \quad (6)$$

где  $q_i$  – знак (-) или (+) для фактора  $q_i$ .

Дисперсия адекватности вычисляется по формуле [1]:

$$D_{\text{Ад}} = \frac{\sum_{j=1}^m (Y_i - \hat{Y}_i)^2}{f} \quad (7)$$

где  $Y_i$  – значение выходного параметра в эксперименте (экспериментальное значение в матрице факторного эксперимента);  $\hat{Y}_i$  – значение выходного параметра, вычисленное по уравнению регрессии (теоретическое значение);  $f$  – число степеней свободы.

Дисперсия воспроизводимости вычисляется по формуле [1]:

$$D_{\text{ВОСПР}} = \frac{\sum_{j=1}^m (Y_i - \bar{Y}_i)^2}{m-1}, \quad (8)$$

где  $\bar{Y}_i$  – среднее значение выходного параметра или свободный член уравнения регрессии  $a_0$ .

После того как вычислены все промежуточные величины, определяются значения  $F_{\text{расч.}}$  и  $dA_i$  по формулам [1]:

$$F_{\text{расч}} = \frac{D_{\text{Ад}}}{D_{\text{ВОСПР}}}, \quad \Delta a_i = \frac{t \cdot \sqrt{D_{\text{ВОСПР}}}}{\sqrt{m}}, \quad (9)$$

где  $t$  – это коэффициент Стьюдента, значение которому присваивается в зависимости от количества строк, полученных при преобразовании матрицы факторного эксперимента, сами значения определены из [1], таблица П.11 (см. табл. 1).

#### 11) Проверка уравнения регрессии на адекватность и сравнение доверительного интервала с коэффициентами уравнения регрессии $dA < a_i$

Все значения критериев Фишера и Стьюдента приведены в [3] и заложены в базу данных программы.

1. Если  $F_{\text{расч.}} > F_{\text{табл.}}$ , то выводим «Уравнение регрессии неадекватно».

2. Если  $F_{\text{расч.}} < F_{\text{табл.}}$ , то продолжаем проверку.

3. Сравниваем все коэффициенты уравнения регрессии с доверительным интервалом.

Здесь выполняется проверка уравнения регрессии на адекватность т. е.  $F_{\text{расч.}} < F_{\text{табл.}}$  [1].

Сравнение доверительного интервала с коэффициентами уравнения регрессии  $dA < a_i$  заключается в следующем: если коэффициент регрессии по модулю меньше доверительного интервала, то этот коэффициент приравниваем к нулю. Здесь программа поочередно сравнивает коэффициенты уравнения регрессии по модулю с доверительным интервалом.

#### 12) Расчет коэффициентов уравнения погрешности и вывод результата на экран

От уравнения регрессии в безразмерном виде переходим к составлению уравнения погрешности.

1. Проводим расчет коэффициентов уравнения погрешности с первого по четвертый.

2. Вывод окончательного результата – уравнения погрешности на экран.

В этой процедуре выполняется расчет конечного результата исследования массива данных по формуле [1]:

$$b_i = \frac{a_i \cdot X_{i0}}{\Delta X_i \cdot a_0}, \quad (10)$$

Уравнение погрешности имеет вид [1]:

$$\frac{\Delta \text{Вых.пар.}}{\text{Вых.пар.}} = b_1 \cdot \frac{\Delta 1 \text{фактор}}{1 \text{фактор}} + b_2 \cdot \frac{\Delta 2 \text{фактор}}{2 \text{фактор}} + b_3 \cdot \frac{\Delta 3 \text{фактор}}{3 \text{фактор}} + b_4 \cdot \frac{\Delta 4 \text{фактор}}{4 \text{фактор}}, \quad (11)$$



где  $\frac{\Delta \text{Вых.пар}}{\text{Вых.пар}}$  – относительное изменение выходного параметра;  $\frac{\Delta 1 \text{фактор}}{1 \text{фактор}}$  – относительное изменение фактора.

Таким образом, нами созданы алгоритм и программа статистической обработки массивов ненормированных данных по методике, предложенной в [2]. С помощью разработанной программы нами были обработаны массивы данных, полученных на примере исследования влияния индивидуальных личностных факторов студентов на результативность обучения по дисциплине «Информатика» [3].

### Ссылки на источники

1. Алексеев В. П., Озёркин Д. В. Основы научных исследований: учеб. пособие для студ. специальности 200800. – Томск: ТУСУР, 2012. – 171 с.
2. Алексеев В. П., Степаньян В. В. Повышение адекватности и достоверности модели обработки ненормированных массивов данных в исследованиях образовательных систем методом планирования эксперимента // Концепт. – 2014. – № 07 (июль). – ART 14178. – URL: <http://e-koncept.ru/2014/14178.htm>.
3. Степаньян В. В. Исследование влияния индивидуальных личностных факторов студентов на результативность обучения по дисциплине «Информатика» // Концепт. – 2014. – № 03 (март). – ART 14070. – URL: <http://e-koncept.ru/2014/14070.htm>.

**Vladimir Stepanian,**

Student, Economics and Technology Department, International Innovative University, Sochi  
[step.wo@mail.ru](mailto:step.wo@mail.ru)

### Algorithms and programs for array processing of continuous data for the educational systems state using experiment planning method

**Abstract.** Studying complex multifactor processes in which the main target parameter is influenced with set of factors, the theory of planning of experiment is applied. Thus values of factors are fixed in both parties from nominal value with the certain interval of a variation. We offer algorithm and the program of processing arrays of factors with any deviations from the nominal value, received during experiment. The original approaches developed earlier are put in a basis. The specified program can be used for optimization of physical, technical, economic and social processes.

**Key words:** algorithm, calculation, coefficient, matrix, function, level of regression, quality, gas.

### References

1. Alekseev, V. P. & Ozjorkin, D. V. (2012) *Osnovy nauchnyh issledovaniy: ucheb. posobie dlja stud. special'nosti 200800*, TUSUR, Tomsk, 171 p. (in Russian).
2. Alekseev, V. P. & Stepan'jan, V. V. (2014) "Povyshenie adekvatnosti i dostovernosti modeli obrabotki nenormirovannyh massivov dannyh v issledovaniyah obrazovatel'nyh sistem metodom planirovaniya jeksperimenta", *Koncept*, № 07 (ijul'), ART 14178. Available at: <http://e-koncept.ru/2014/14178.htm> (in Russian).
3. Stepan'jan, V. V. (2014) "Issledovanie vlijaniya individual'nyh lichnostnyh faktorov studentov na rezul'tativnost' obuchenija po discipline 'Informatika'", *Koncept*, № 03 (mart), ART 14070. Available at: <http://e-koncept.ru/2014/14070.htm> (in Russian).

### Рекомендовано к публикации:

Довгалем Л. С., кандидатом педагогических наук, профессором;

Горевым П. М., кандидатом педагогических наук,

главным редактором журнала «Концепт»



[www.e-koncept.ru](http://www.e-koncept.ru)

Поступила в редакцию <i>Received</i>	30.01.15	Получена положительная рецензия <i>Received a positive review</i>	02.02.15
Принята к публикации <i>Accepted for publication</i>	02.02.15	Опубликована <i>Published</i>	30.04.15

© Концепт, научно-методический электронный журнал, 2015

© Степаньян В. В., 2015