

УДК 658.5

UDC 658.5

РАНЖИРОВАНИЯ ПЕРЕМЕННЫХ ДЛЯ ИДЕНТИФИКАЦИИ ЦЕЛЕВЫХ ФУНКЦИЙ В ЗАДАЧАХ УПРАВЛЕНИЯ И АВТОМАТИЗАЦИИ ПРОЕКТИРОВАНИЯ

RANGINGS OF VARIABLES FOR IDENTIFICATION OF CRITERION FUNCTIONS IN PROBLEMS OF MANAGEMENT AND DESIGNING AUTOMATION

Стародубцев Виктор Сергеевич
д.т.н., к.г.-м.н., профессор
Российский государственный социальный университет филиал в г.Воронеж, Воронеж, Россия

Starodubtsev Victor Sergeevich
Dr.Tech.Sci., Cand.Geol.-Min.Sci., professor
Russian state social university branch in Voronezh, Voronezh, Russia

Шхачева Рита Гучипсовна
Майкопский государственный технологический университет, Майкоп, Россия

Shkhacheva Rita Guchipsova
Maykop state technological university, Maykop, Russia

Рассматривается алгоритм выбора независимых параметров в многопараметрических задачах идентификации целевых функций производства

The algorithm of a choice of independent parameters in multiple parameter problems of identification of criterion functions of manufacture is considered

Ключевые слова: АЛГОРИТМ, СТРУКТУРНАЯ ИДЕНТИФИКАЦИЯ, СИСТЕМА, СИСТЕМЫ АВТОМАТИЗАЦИИ ПРОЕКТИРОВАНИЯ

Keywords: ALGORITHM, STRUCTURAL IDENTIFICATION, SYSTEM, SYSTEMS OF AUTOMATION OF DESIGNING

Исследуя и моделируя сложные объекты, мы в большинстве случаев не можем точно определить, какие из параметров влияют на исследуемую систему, а какие есть производные первых. Для выяснения этого к перебору допускается как можно больше известных нам параметров, чтобы исключить один из моментов субъективного подхода в моделировании [1, 3]. С другой стороны, увеличивая количество исходных членов полного описания, мы увеличиваем время решения задачи, а также возникает возможность появлению зависимых переменных, что в свою очередь влечет плохую обусловленность матрицы коэффициентов и приводит в конечном результате к неверным решениям. Определение коэффициента парной корреляции

$$R_{i,j} = \begin{cases} \frac{\sum_{\gamma=M_1}^{M_2} X_{\gamma,i} X_{\gamma,j}}{M_2 - M_1 + 1}, & j \neq i; \\ 1 & j = i, \end{cases} \quad (1)$$

где M_1 - начало выборки, а M_2 - конец выборки, $i, j = 1, 2, \dots, n$, где n – количество переменных взятых для проведения эксперимента по идентификации целевых функций изучаемых процессов.

Коэффициент парной корреляции позволяет исключить из процесса идентификации модели с зависимыми переменными, т.е. с коэффициентом $R_0 > 0,7$. При большом количестве входных параметров процесс отбора независимых параметров становится неэлементарной задачей.

Для обеспечения автоматизированного корректного выбора независимых параметров во множестве фактических данных предлагается алгоритм ранжирования переменных. Алгоритм основан на ранжировании переменных (таблица 1), исходя из значения соответствующих коэффициентов парной корреляции между отдельными переменными.

Таблица 1 - РАНЖИРОВАНИЕ ПЕРЕМЕННЫХ ПО ЗНАЧЕНИЮ ВЕЛИЧИНЫ КОЭФФИЦИЕНТА ПАРНОЙ КОРРЕЛЯЦИИ $R_{i,j}$

| Значение коэффициента парной корреляции | Ранг |
|---|------|
| $R_{i,j} > 0,9$ | 1,25 |
| $0,9 < R_{i,j} < 0,85$ | 1 |
| $0,85 < R_{i,j} < 0,8$ | 0,75 |
| $0,8 < R_{i,j} < 0,75$ | 0,5 |
| $0,75 < R_{i,j} < 0,7$ | 0,25 |

Суммарный ранг переменной x_i определяется в соответствии с выражением

$$RANG_i = \sum_{j=1}^{n-1} R_{i,j} . \quad (2)$$

Из эксперимента исключаются переменные с максимальными значениями $RANG$. Количество переменных допускаемых для идентификации модели изучаемого процесса (n^*) определяется исходя из возможности метода и (или) исходя из исключения всех переменных с $R_{i,j} > 0,7$. В общем виде структурную схему алгоритма ранжирования переменных можно представить в виде рисунке 1.

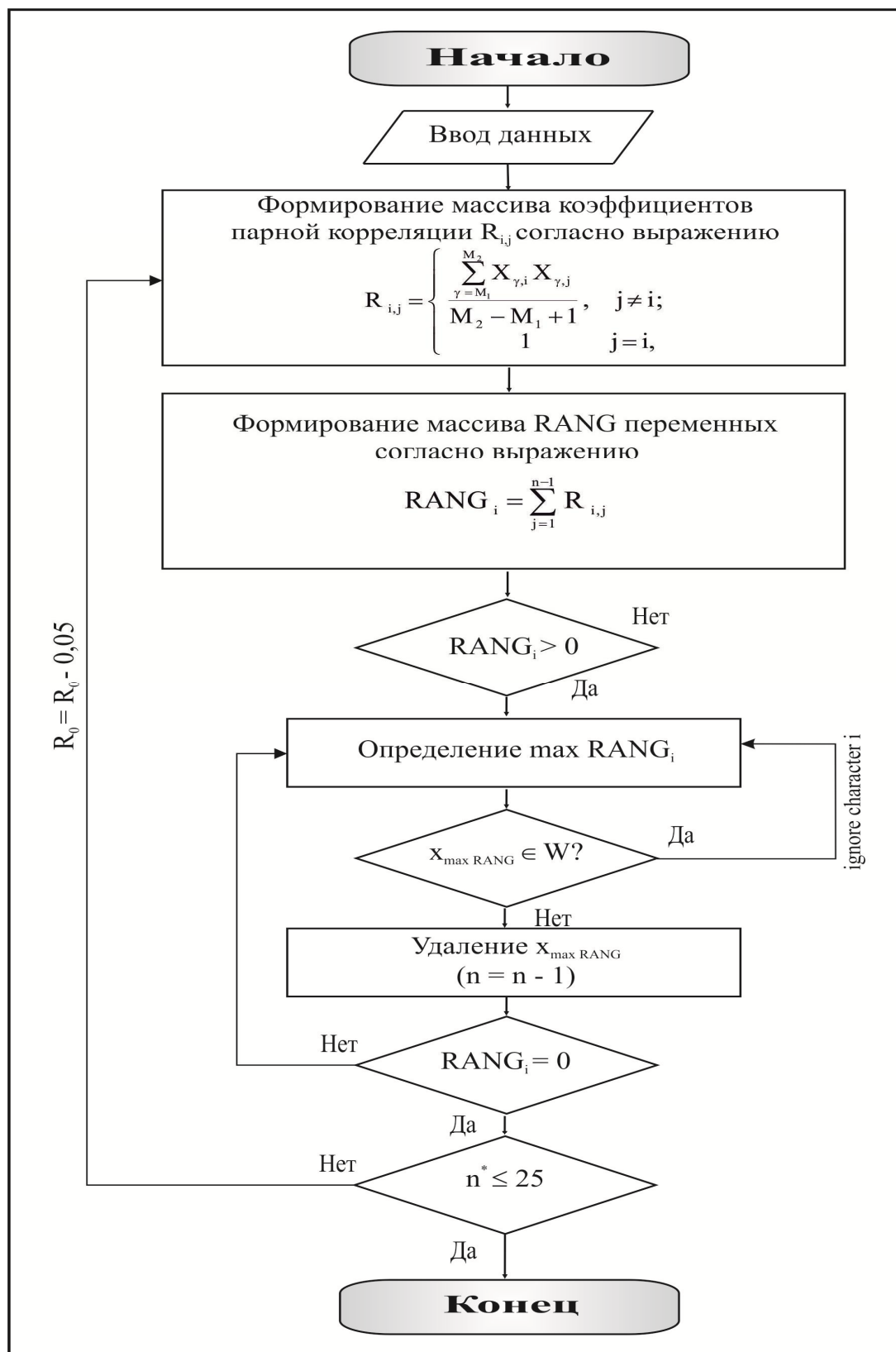


Рисунок 1. Алгоритм ранжирования переменных по величине коэффициента парной корреляции (W – множество управляющих параметров)

Предлагаемый алгоритм ранжирования переменных был использован для идентификации целевых функций в задачах автоматизированного проектирования систем водозаборов подземных вод с учетом качественной составляющей питьевой воды и поддержки принятия управленческих решений с использованием метода группового учета аргументов (МГУА (Group Method of Data Handling, GMDH)).

Задача математического моделирования по выборкам наблюдений за протеканием исследуемого процесса - это задача восстановления зависимости вида

$$y = h(x) + \eta, \quad h: R^m \rightarrow R^n, \quad (3)$$

где $\eta = (\eta_1, \dots, \eta_n)$ - ошибка наблюдения.

Задачу восстановления зависимости (3), решаемую с помощью МГУА, можно сформулировать следующим образом [1]. Пусть выполнены следующие предположения:

- существует единственная зависимость $h: R^m \rightarrow R^n$ между входной X и выходной Y величинами;
- задан класс F – класс структур отображений;
- задана $(n \times m)$ матрица X значений входной величины (выборка):
- $Y = (y_0^1 + \eta_1, \dots, y_0^n + \eta_n)$, где y_0^1, \dots, y_0^n - истинные значения выходной величины в n точках, $\eta = (\eta_1, \dots, \eta_n)$ - ошибка наблюдений, причем величины η_i , $(i=1, \dots, n)$ предполагаются случайными, независимыми, одинаково распределенными с нулевым математическим ожиданием и конечной дисперсией.

В этих предположениях требуется решить задачу нахождения такой структуры $f^* \in F$, что

$$s(f^*) = \inf_{f \in F} S(f)$$

или (если $f^* \notin F$) такой структуры $f^e \in F$, что

$$s(f^e) = \inf_{f \in F} S(f) + \varepsilon, \quad \varepsilon > 0.$$

Учитывая один из основных принципов теории МГУА - принцип "свободы выбора решений", дифференциальное уравнение конвективно-диффузионного переноса мигранта в двумерном потоке [2] для задачи автоматизированного проектирования систем водозаборов подземных вод с учетом качественной составляющей питьевой воды используем полное описание класса структур для идентификации процесса массопереноса загрязняющих компонентов подземных вод вида

$$\frac{\partial c}{\partial t} = a_1 \left(\frac{\partial^2 c}{\partial x^2} \right)^{t-k} + a_2 \left(\frac{\partial^2 c}{\partial y^2} \right)^{t-k} + a_3 \left(\frac{\partial c}{\partial x} \right)^{t-k} + a_4 \left(\frac{\partial c}{\partial y} \right)^{t-k} + a_5 \Psi_1^{t-k} + a_6 \Psi_2^{t-k} + a_7 \Psi_3^{t-k} + a_8 \Psi_4^{t-k} + a_9 \Psi_5^{t-k} + a_{10} \Psi_6^{t-k} + a_{11}, \quad (4)$$

где c – концентрация ионов загрязняющего компонента (например, железа или марганца) в подземных водах (прогнозируемая переменная в мг/л); $a_1 - a_4$ - соответствующие коэффициенты при производных; t - время; x, y - пространственные координаты; k – запаздывание по времени, $k = 1, 2, 3$; Ψ_1 - водоотбор в тыс.м³/сут; Ψ_2 - температура воздуха в С⁰; Ψ_3 – осадки в мм; Ψ_4 - рН поверхностных вод; Ψ_5 - содержание О₂ в поверхностных водах водохранилища мг/л; Ψ_6 - содержание ионов хлора в поверхностных водах водохранилища мг/л; $a_5 - a_{10}$ - соответствующие коэффициенты при $\Psi_1 - \Psi_6$; a_{11} - свободный член.

Для задачи маркетинга, как объекта математического моделирования, по реализации прогнозирования экономической деятельности предприятий промышленности рассматривались на основе следующего класса структур

$$\frac{\partial Y}{\partial t} = a_1 x_1^{t-m} + a_2 x_2^{t-m} + \mathbf{K} + a_k v_k^{t-m} + a_{k+1} v_{k+1}^{t-m} + \mathbf{K} + a_{n-1} w_{n-1}^{t-m} + a_n w_n^{t-m} + a_0, \quad (5)$$

где Y – идентифицируемая функция, x_1, x_2, \dots – внешние параметры производства, v_k, v_{k+1}, \dots - внутренние параметры, w_{n-1}, w_n – управляющие параметры, t – время, $a_1 - a_n$ – коэффициенты при параметрах социально-

экономической системы предприятия, a_0 – свободный член, $m = 0, 1, 2$ – шаг запаздывания, который равен кварталу.

Таким образом, на базе полного описания (4) рассматривались 10 параметров, а с учетом запаздывания по времени (0, 1 и 2) – 30 параметров.

В задаче маркетинга (5) количество параметров определяется возможностями исследователя провести мониторинг изучаемой социально-экономической системы предприятия. Как правило, количество параметров в этом случае может достигать несколько десятков и даже сотен.

Для комбинаторного алгоритма МГУА [1] $\max n^* = 25$, т.е. максимально может осуществляться перебор не более чем $2^{25} - 1$ моделей претендентов. Рассмотрим возможности предлагаемого алгоритма для задачи маркетинга (5).

В задаче маркетинга для пищевого предприятия рассматривалась возможность повышения рентабельности производства. В этом случае проводились исследования по идентификации целевой функции формирования выручки от продажи товаров (Y_1) и основных составляющих себестоимости продукции цены на молоко (Y_2), масло (Y_3), муку (Y_4) и сахар (Y_5). Для идентификации процесса формирования выручки от продажи товаров (Y_1) на базе полного описания (3) было использовано 46 параметров, а с учетом запаздывания по времени (0, 1 и 2) – 138 параметров. Аналогично для процессов формирования цен на молоко (Y_2), масло (Y_3), муку (Y_4) и сахар (Y_5) – 22 независимых параметров и с учетом запаздывания по времени (0, 1 и 2) – 66 независимых параметров.

Использование 138 независимых параметров для процесса идентификации модели "Выручка от продажи товаров" и 66 независимых параметров для моделей процессов формирования цен на молоко, масло, муку и сахар потребовало использования алгоритма ранжирования переменных при идентификации целевых функций рассматриваемых

процессов. Оценка эффективности отбора проводилась по критерию эпигнозного прогноза (P) (рис. 2 - 6).

$$P = \sqrt{\frac{\sum_{i=D+1}^C (Y_{im_D} - Y_{it})^2}{\sum_{i=D+1}^C Y_{it}^2}}, \quad (6)$$

где Y_{im_D} – модель, полученная по выборке D , Y_{it} – табличное значение выходной величины, при этом мощность выборки фактических данных Ω рассматривалась в виде $\Omega = D+C$, где $D=A+B$.

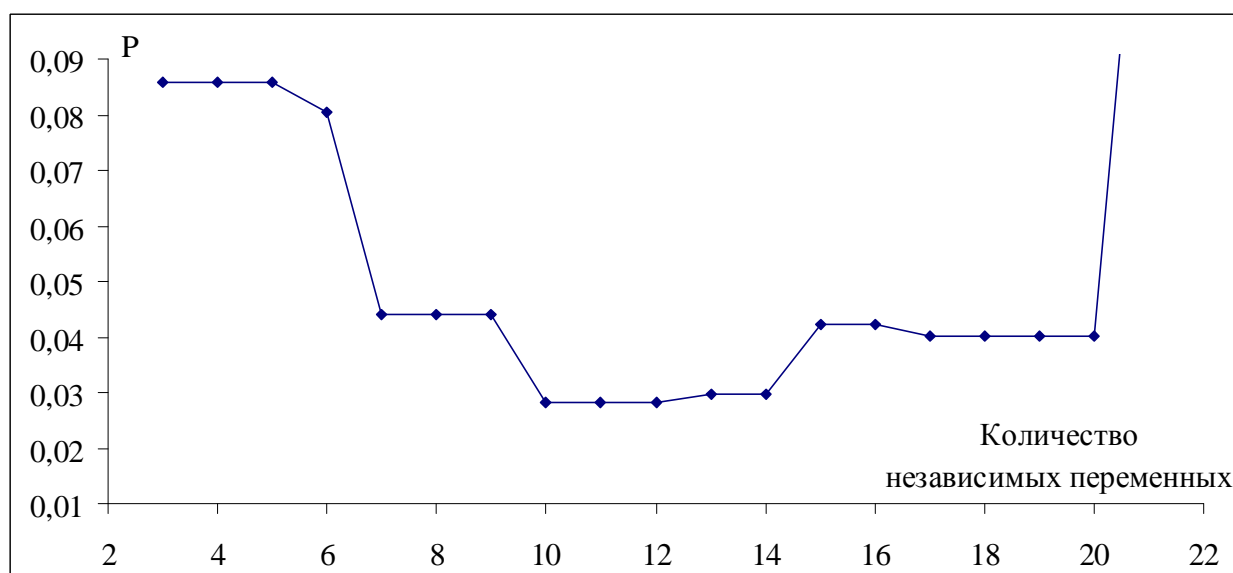


Рисунок 2. Значение критерия эпигнозного прогноза (P) при изменении числа независимых параметров для параметра "Выручка от продажи товаров"

Результаты оптимизации числа независимых параметров сведены в таблицу 2.

Интервал 10-12 независимых параметров для модели "Выручка от продажи товаров" объясняется тем, что при изменении числа независимых параметров выбиралась одна и та же модель.

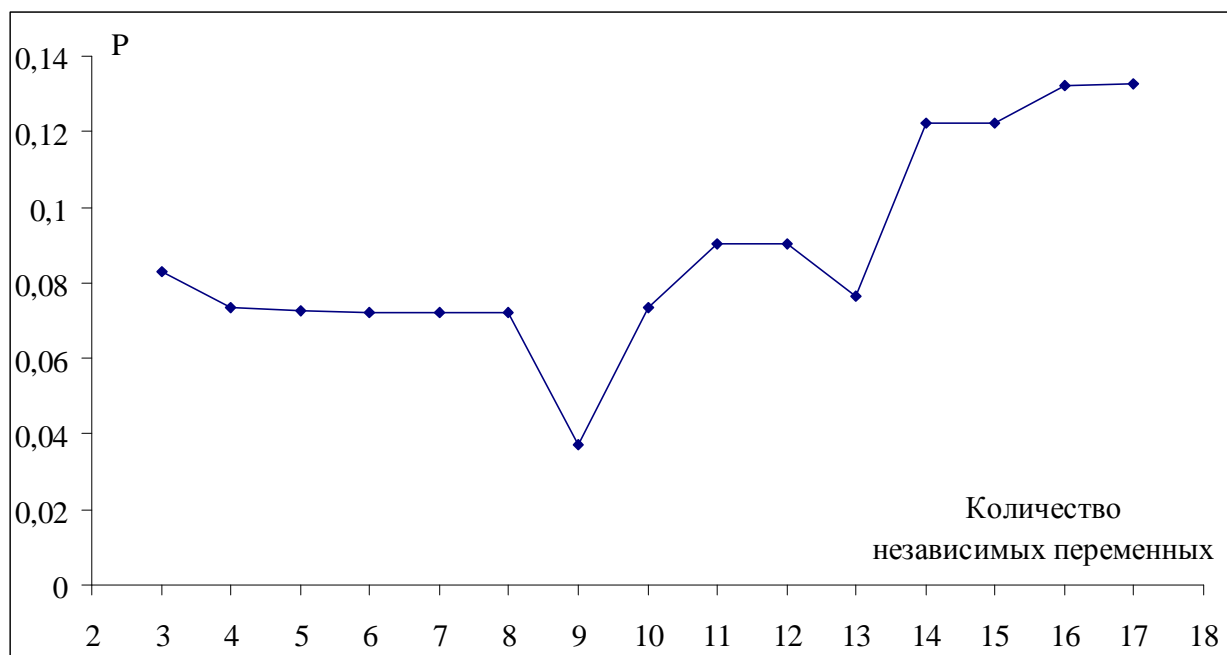


Рисунок 3. Значение критерия эпигнозного прогноза (P) при изменении числа независимых параметров для параметра "Молоко"

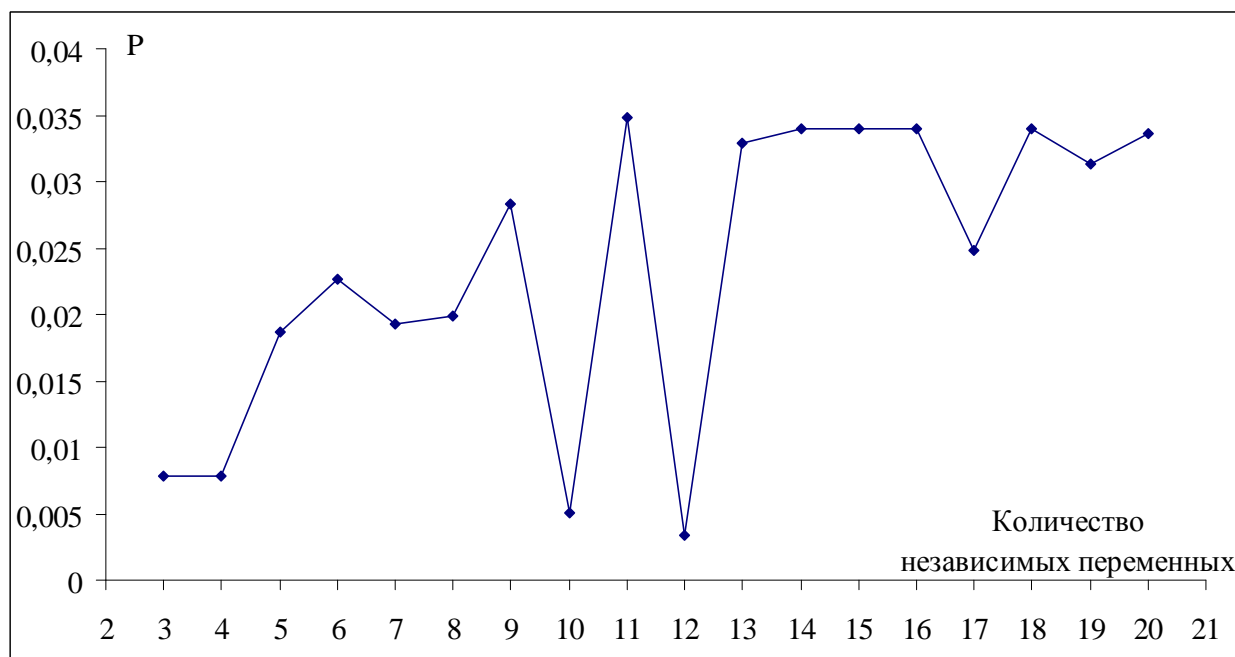


Рисунок 4. Значение критерия эпигнозного прогноза (P) при изменении числа независимых параметров для параметра "Масло"

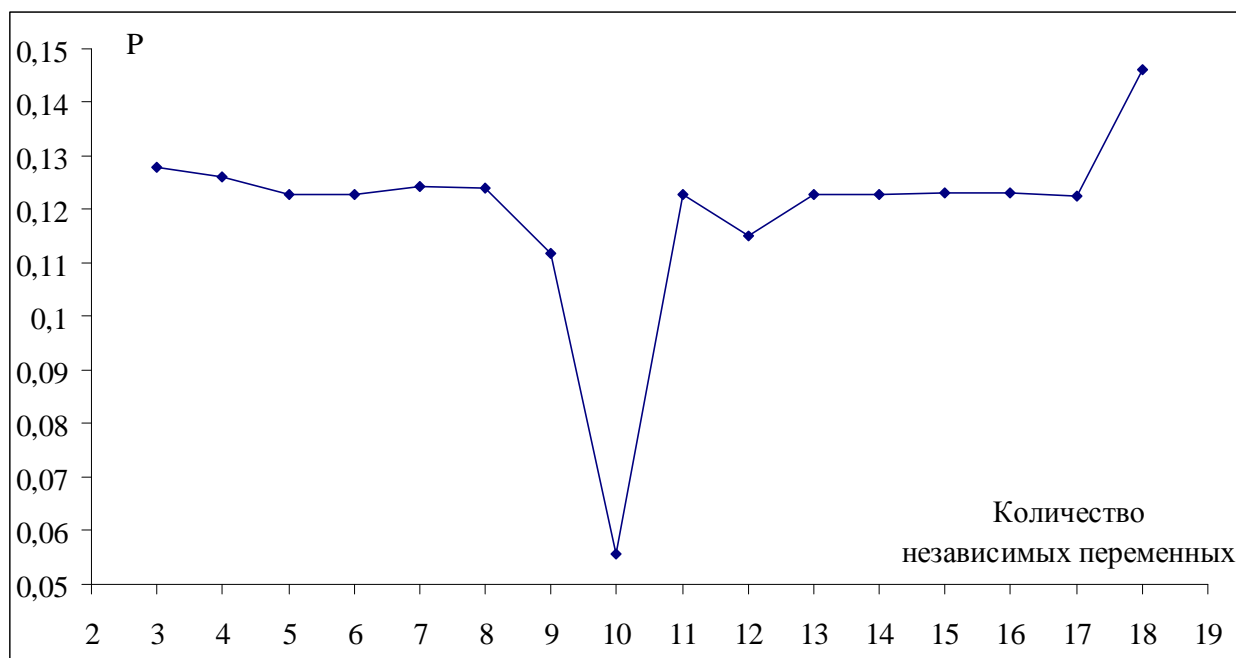


Рисунок 5. Значение критерия эпигнозного прогноза (P) при изменении числа независимых параметров для параметра "Мука"

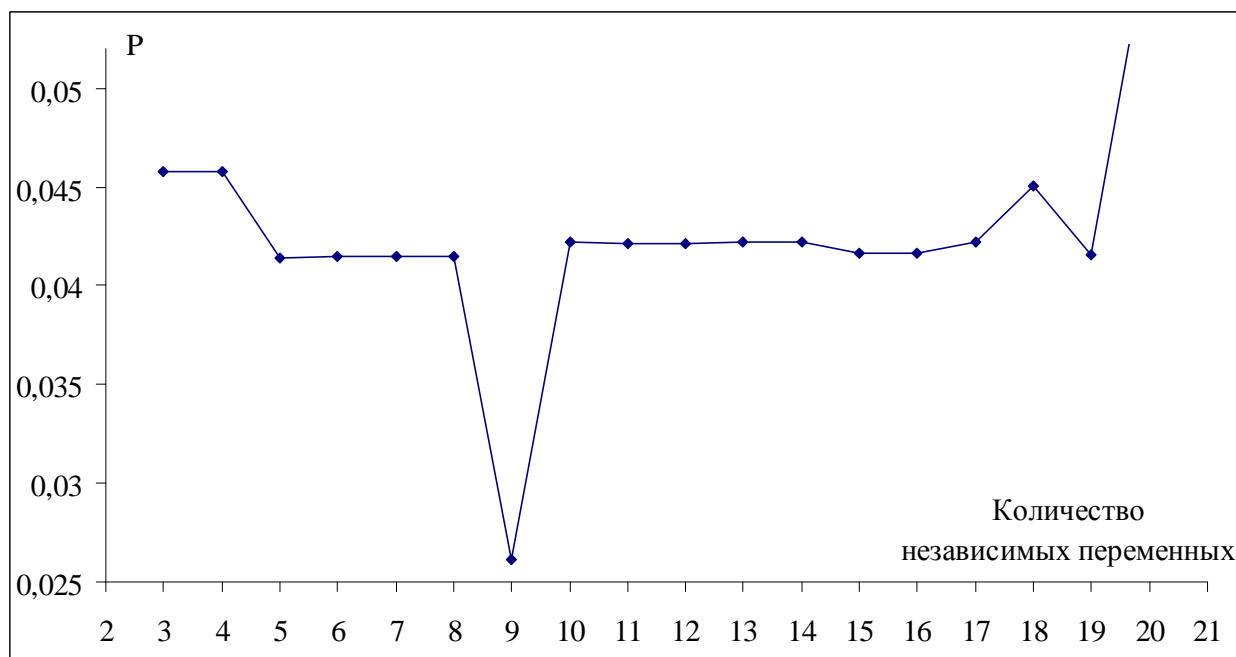


Рисунок 6. Значение критерия эпигнозного прогноза (P) при изменении числа независимых параметров для параметра "Сахар"

Таблица 2 – ЧИСЛО НЕЗАВИСИМЫХ ПАРАМЕТРОВ, ОТОБРАННЫХ
ДЛЯ ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТОВ ПО ИДЕНТИФИКАЦИИ
ЭКОНОМИКО-СОЦИАЛЬНЫХ ПРОЦЕССОВ ПИЩЕВОГО
ПРОИЗВОДСТВА

| Параметр | Значение критерия эпигнозного прогноза (P) | Число независимых параметров |
|-------------------------------|---|---------------------------------|
| Выручка от продажи товаров | 0,02824 | 10-12 |
| Молоко | 0,03704 | 9 |
| Масло | 0,00336 | 12 |
| Мука | 0,05577 | 10 |
| Сахар | 0,02614 | 9 |

Полученные результаты эксперимента по идентификации моделей экономических процессов пищевого производства позволяют сделать вывод об эффективности предложенного алгоритма ранжирования переменных при идентификации целевых функций, позволяющего значительно сократить время проведения эксперимента (что определяет количество допущенных к эксперименту переменных n^*) и получать прогностические модели экономических процессов для целей управления, в том числе и для принятия управленческих решений в связи с малой погрешностью прогноза (3% - 5%).

Литература

1. Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем.–Киев: Наук. думка, 1982.–296 с.
2. Шестаков В.М. Динамика подземных вод.-М.: Изд-во МГУ, 1979.-368 с.
3. Шхачева, Р.Г. Моделирование бизнес-процессов с учетом затрат на маркетинг // Системы управления и информационные технологии. 2010. №1.2.(39). С. 280- 284.