

## ПРИМЕНЕНИЕ ДИСПЕРСИОННОГО АНАЛИЗА ДЛЯ ОПТИМИЗАЦИИ КОЛИЧЕСТВА ДАТЧИКОВ ОХРАННОЙ ЗОНЫ ПРОДУКТОПРОВОДА

Для защиты продуктопровода от несанкционированных врезок с целью хищения перекачиваемого агента (нефть, газ и пр.) применяются автоматизированные системы мониторинга несанкционированных действий в охранной зоне продуктопровода. В качестве датчиков обнаружения злоумышленников используются геофоны. С помощью метода дисперсионного анализа предложено оптимальное количество датчиков на разные участки продуктопроводов.

**Ключевые слова:** дисперсионный анализ, несанкционированные действия, врезки, геофоны, автоматизированная система, продуктопровод.

Транспорт нефти и газ на территории РФ осуществляется с помощью многокилометровых систем продуктопроводов. Для защиты данных систем транспорта от несанкционированных врезок со стороны злоумышленников с целью хищения перекачиваемого агента предусматриваются различные технологии и технические средства защиты. Одно из таких решений защиты – автоматизированная система мониторинга несанкционированных действий в охранной зоне продуктопровода (АСМНД).

Принцип работы системы [1]:

Несанкционированные действия сопровождаются механическими воздействиями, производимыми человеком или техникой на грунт в охранной зоне продуктопровода. Данные воздействия создают шум, который улавливают геофоны (датчики), расположенные на охраняемом участке. Группа геофонов располагается на охраняемом участке продуктопровода. Информация с геофонов в виде нормируемого сигнала передаётся на шкаф сбора данных с микропроцессорными устройствами. Микропроцессорные устройства фиксируют и накапливают информацию с геофонов. Далее по заложенным алгоритмам в шкафу сбора данных происходит анализ параметров. Обработанные данные с микропроцессорных устройств передаются на уровень оперативно-производственной службы, на АРМ диспетчера.

На рисунке 1 приведена структурная схема регистрации сигнала с геофонов.

Экспериментальными данными установлено следующее: для охраны участка длиной 10 км от несанкционированных врезок в системе необходимо использовать 9–11 датчиков. Однако существуют участки меньшей длины. Следовательно,

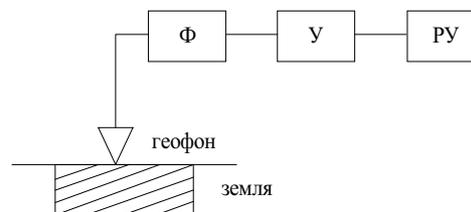
но, для таких участков необходимо меньшее количество датчиков. Рассмотрим предположение, что минимальное количество датчиков равно 3 шт. на 1 км. Минимальное количество выбрано с учётом возможных погрешностей и вероятностью выхода из строя одного из датчиков.

Рассмотрим несколько продуктопроводов с разными длинами участков и разным количеством датчиков на этих участках. Для проверки нашего утверждения о минимальном количестве датчиков на различных участках продуктопроводах впервые применен метод дисперсионного анализа.

Дисперсионный анализ был разработан и предложен английским ученым, математиком и генетиком Рональдом Фишером в 20-х годах XX века.

Дисперсионный анализ – это статистический метод оценки связи между факторными и результативным признаками в различных группах, отобранный случайным образом, основанный на определении различий (разнообразия) значений признаков.

Дисперсионный анализ – метод, проверки гипотез о равенстве трёх и более средних, основанный на F-критерии [2].



Ф – фильтр, У – усилитель, РУ – регистрирующее устройство

Рисунок 1. Структурная схема регистрации сигнала с геофона

Анализ, который рассматривает только одну переменную, называют однофакторным дисперсионным анализом (метод, проверяющий влияние на зависимую переменную одной независимой переменной (фактора)). В нашем случае исследуемая зависимая переменная – количество датчиков. Рассматривается только один воздействующий фактор – участок продуктопровода (участок 1,2,3,4), который имеет 5 различных интервалов.

В основе дисперсионного анализа лежит анализ отклонений всех единиц исследуемой совокупности от среднего арифметического. В качестве меры отклонений берется дисперсия – средний квадрат отклонений. Отклонения, вызываемые воздействием факторного признака (фактора) сравниваются с величиной отклонений, вызываемых случайными обстоятельствами. Если отклонения, вызываемые факторным признаком, более существенны, чем случайные отклонения, то считается, что фактор оказывает существенное влияние на результативный признак.

Факторные признаки – это те признаки, которые влияют на изучаемое явление. Результативные признаки – это те признаки, которые изменяются под влиянием факторных признаков.

Для расчёта используем Метод по Фишеру (Fisher) с критерий F (значения F выбираются с помощью теоретических значений распределения Фишера-Снедекора). Метод применяется в однофакторном дисперсионном анализе, когда совокупная дисперсия всех наблюдаемых значений раскладывается на дисперсию внутри отдельных групп и дисперсию между группами.

Принцип применения метода дисперсионного анализа

Формулируем нулевую гипотезу ( $H_0$ ), то есть, предполагаем, что исследуемые факторы (количество датчиков) не оказывают никакого влияния на значения результативного признака (длины участков) и полученные различия случайны.

Затем определяем, какова вероятность получить наблюдаемые (или более сильные) различия при условии справедливости нулевой гипотезы.

Если эта вероятность мала, то мы отвергаем нулевую гипотезу и заключаем, что результаты исследования статистически значимы.

Алгоритм расчёта:

1. Построение дисперсионного комплекса.
2. Вычисление средних квадратов отклонений.
3. Вычисление дисперсии.
4. Сравнение факторной и остаточной дисперсий.
5. Оценка результатов с помощью теоретических значений распределения Фишера-Снедекора.

### 1. Построение дисперсионного комплекса.

Построим таблицу 1 с результативными и факторными признаками

### 2. Вычисление средних квадратов отклонений.

Вычисление суммы вариант по каждой градации факторов осуществляется по следующей формуле:

$$\sum V_j = V_1 + V_2 + V_3$$

Вычисление общей суммы вариант ( $\sum V_{\text{общ.}}$ ) по всем градациям факторного признака осуществляется по следующей формуле:

$$\sum V_{\text{общ.}} = \sum V_{j1} + \sum V_{j2} + \sum V_{j3}$$

Расчёт средних квадратов отклонений представлен в таблице 2.

### 3. Вычисление дисперсий.

Вычисление дисперсии:

$$D_{\text{общ.}} = D_{\text{факт.}} + D_{\text{ост.}}$$

где  $D_{\text{общ.}}$  – общая дисперсия, характеризуется разбросом вариант (наблюдаемых значений) от общего среднего;

$D_{\text{факт.}}$  – факторная (межгрупповая) дисперсия, характеризует разброс групповых средних от общего среднего;

$D_{\text{ост.}}$  – остаточная (внутригрупповая) дисперсия, характеризует рассеяние вариант внутри групп.

Вычисление факториальной дисперсии:

$$D_{\text{факт.}} = \sum h - H$$

Вычисление h проводится по формуле:

$$h = \frac{(\sum V_j)^2}{n};$$

Вычисление H проводится по формуле:

$$H = \frac{(\sum V)^2}{\sum n};$$

Вычисление остаточной дисперсии:

$$D_{ост.} = \sum V_{общ.}^2 - \sum h;$$

Вычисление общей дисперсии:

$$D_{общ.} = \sum V_{общ.}^2 - H$$

Расчёт дисперсий представлен в таблице 3.

**4. Расчет основного показателя силы влияния изучаемого фактора.**

Показатель силы влияния  $\eta^2$  факторного признака на результат определяется долей фак-

ториальной дисперсии  $D_{факт.}$  в общей дисперсии  $D_{общ.}$ ,  $\eta^2$  (эта) – показывает какую долю, занимает влияние изучаемого фактора среди всех других факторов и определяется по формуле:

$$\eta^2 = \frac{D_{факт.}}{D_{общ.}} * 100\%$$

Расчёт основного показателя представлен в таблице 4.

Таблица 1. Результативные и факторные признаки

Участок продуктопровода	Количество геофонов (факторный признак)				
	группировки фактора				
	до 1 км	от 1 до 2 км	от 2 до 4 км	от 4 до 8 км	от 8 до 10 км
участок 1	3	3	3	5	9
участок 2	3	3	4	6	10
участок 3	3	3	4	7	11
участок 4	3	3	4	8	11

Таблица 2. Расчёт средних квадратов отклонений

Применяемые формулы и условные обозначения	Градации факторного признака r=4 (количество градаций)				
	3,3,3,3	3,3,3,3	3,4,4,4	5,6,7,8	9,10,11,11
V	3,3,3,3	3,3,3,3	3,4,4,4	5,6,7,8	9,10,11,11
N	4	4	4	4	4
$\sum V_j$	3+3+3+3=12	3+3+3+3=12	3+4+4+4=15	5+6+7+8=26	9+10+11+11=41
$\sum V_{общ.}$	12+12+15+26+41=106				
$\sum (V^2) = V_1^2 + V_2^2 + V_3^2$	9+9+9+9=36	9+9+9+9=36	9+16+16+16=57	25+36+49+64=174	81+100+121+121=423
$\sum V^2_{общ.}$	36+36+57+174+423=726				
$(\sum V_j)^2$	144	144	225	676	1681

Таблица 3. Расчёт дисперсий

Применяемые формулы и условные обозначения	Градации факторного признака r=4 (количество градаций)				
	144/4=36	144/4=36	225/4=56,25	676/4=169	1681/4=420,25
$h = \frac{(\sum V_j)^2}{n}$	144/4=36	144/4=36	225/4=56,25	676/4=169	1681/4=420,25
$\sum h$	36+36+56,25+169+420,25=717,45				
$H = \frac{(\sum V)^2}{\sum n}$	11236/20=561,8				
$D_{факт.} = \sum h - H$	717,45-561,8=155,65				
$D_{общ.} = \sum V^2_{общ.} - H$	726-561,8=164,2				
$D_{ост.} = \sum V^2_{общ.} - \sum h$	726-717,45=8,55				

Таблица 4. Расчет основного показателя силы влияния изучаемого фактора

Применяемые формулы и условные обозначения	Градации факторного признака $r=4$ (количество градаций)
$\eta^2 = \frac{D_{\text{факт.}}}{D_{\text{общ.}}} * 100\%$	155,65/164,2=94,79
$\sigma^2_{\text{факт.}} = \frac{D_{\text{факт.}}}{r-1}$	155,65/3=51,88
$\sigma^2_{\text{ост.}} = \frac{D_{\text{ост.}}}{N-r}$	8,55/15=0,57
$F = \frac{\sigma^2_{\text{факт.}}}{\sigma^2_{\text{ост.}}}$	51,88/0,57=91,01
степени свободы $v1 = n - 1$ $v2 = N - 1$	$v1 = 4 - 1 = 3$ $v2 = 20 - 5 = 15$
$F \geq F_{st.}$	91,01 > от 3,3 до 5,4

### 5. Оценка результатов с помощью теоретических значений распределения Фишера-Снедекора

$$F = \frac{\sigma^2_{\text{факт.}}}{\sigma^2_{\text{ост.}}} \geq F_{st.},$$

где  $F$  – критерий Фишера;

$F_{st.}$  – табличное значение [3].

$\sigma^2_{\text{факт.}}$  и  $\sigma^2_{\text{ост.}}$  – факториальная и остаточная дисперсия отклонение от средней линии, определяются по формулам:

$$\sigma^2_{\text{ост.}} = \frac{D_{\text{ост.}}}{N-r}$$

где  $r$  – число градаций факторного признака.

Сравнение критерия Фишера ( $F$ ) со стандартным (табличным)  $F$  проводим по графам таблицы с учетом степеней свободы:

$$v1 = n - 1$$

$$v2 = N - 1$$

По горизонтали определяем  $v1$  по вертикали  $v2$ , на их пересечении определяют табличное значение  $F$ , где верхнее табличное значение  $p \geq 0,05$ , а нижнее соответствует  $p \geq 0,01$ , и сравнивают с вычисленным критерием  $F$ . Если значение вычисленного критерия  $F$  равно или больше табличного, то результаты достоверны и  $H_0$  не отвергается [4].

### Вывод

Выявлено, что сила влияния количества датчиков на участках продуктопровода составляет 94,79%. Для всех участков можно с вероятностью 91,01 утверждать, что выбранное количество датчиков соответствует длинам участков. Таким образом, нулевая гипотеза ( $H_0$ ) не отвергается и предложенное количество датчиков на разные длины участков считается доказанным.

28.10.2013

Работа выполнена при поддержке Российского фонда фундаментальных исследований (Договор №НР 13-07-97025/13 от 10.07.2013 г.), а также Министерства образования Оренбургской области (Соглашение №13-07-97025 от 23.07.2013 г.)

### Список литературы:

1. Кораблев Е.Н., Ягудина Л.В., Сосновцева Е.В. Решения по структуре и функционированию автоматизированной системы мониторинга несанкционированных действий в охранной зоне продуктопровода, требования к основным элементам системы // Проблемы сбора, подготовки и транспорта нефти и нефтепродуктов. – 2012. – №4. – С. 137–142.
2. Иванов О.В. Статистика. Учебный курс для социологов и менеджеров. Часть 2. Доверительные интервалы. Проверка гипотез. Методы и их применение. – М. 2005. – 220 с.

3. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 592 с.
4. Шеффе Г. Дисперсионный анализ. – М.: Наука. Главная редакция физико-математической литературы, 1980 – С. 476–485.

Сведения об авторах:

**Кораблев Евгений Николаевич**, магистрант кафедры алгебры и математической кибернетики математического факультета Оренбургского государственного университета  
460018, г. Оренбург, пр-т Победы, 13, ауд. 1410, тел. (3532) 276976, e-mail: liss-evgenii@mail.ru  
**Влацкая Ирина Валерьевна**, заведующая кафедрой компьютерной безопасности и математического обеспечения информационных систем математического факультета  
Оренбургского государственного университета, кандидат технических наук, доцент  
460018, г. Оренбург, пр-т Победы, 13 ауд. 2131, тел. (3532) 372534, e-mail: mois@mail.osu.ru

**UDC 681.5:658.382**

**Korablev E.N. Vlatskaya I.V.**

Orenburg state university, e-mail: liss-evgenii@mail.ru

**APPLICATION OF THE DISPERSIVE ANALYSIS FOR OPTIMIZATION OF NUMBER OF SENSORS SECURITY ZONE OF THE PRODUCTS PIPELINE.**

For the purpose of plunder of the pumped-over agent (oil, gas and so forth) are applied to protection of the products pipeline against unauthorized inserts the automated systems of monitoring of unauthorized actions in a security zone of the products pipeline. As sensors of detection of malefactors geophones are used. By means of a method of the dispersive analysis the optimum quantity the sensor on different sites of products pipelines is offered.

Key words: the dispersive analysis, unauthorized actions, inserts, the geophones, the automated system, the products pipeline.

**Bibliography:**

1. Korablev E.N. Yagudin L.V. Sosnovtseva E.V. Decisions on structure and functioning of the automated system of monitoring of unauthorized actions in a security zone of the products pipeline, the requirement to basic elements of system// Problems of collecting, preparation and transport of oil and oil products. – 2012. – No. 4. – Page 137-142.
2. Ivanov O. V. Statistics. Training course for sociologists and managers. Part 2. Confidential intervals. Check of hypotheses. Methods and their application. – M 2005. – 220 pages.
3. Kobzar A.I. Prikladnaya mathematical statistics. For engineers and scientists. – M: FIZMATLIT, 2006. – 592 pages.
4. Scheff G. Dispersive analysis. – M: Science. The main edition of physical and mathematical literature, 1980 – Page 476-485.