

## ОЦЕНКА СТАТИСТИЧЕСКОЙ ПОДКОНТРОЛЬНОСТИ В МЕДИКО-ЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

<sup>1</sup>С.А.Лаптёнок, <sup>1</sup>О.И.Родькин, <sup>2</sup>А.А.Кологривко, <sup>2</sup>С.А.Хорева, <sup>2</sup>Ю.В.Кляусова

<sup>1</sup>Международный государственный экологический институт им. А.Д.Сахарова  
Белорусского государственного университета,  
ул. Долгобродская, 23/1, г. Минск, 220070, г. Минск, Республика Беларусь  
<sup>2</sup>Белорусский национальный технический университет,  
пр. Независимости, 65, 220013, г. Минск, Республика Беларусь

*Представлены результаты оценки эффективности оценки статистической подконтрольности данных в медико-экологических исследованиях с использованием метода восходящих и нисходящих серий.*

*Ключевые слова: статистическая подконтрольность; метод восходящих и нисходящих серий; медико-экологические исследования.*

Изучаемое явление подконтрольно, если можно предсказать его будущее течение. Соответственно этому явление считается статистически подконтрольным, если предсказание возможно в вероятностном смысле, то есть, если можно предсказать вероятность попадания явления в определенные пределы. Значит, результаты случайной операции будут статистически контролируемыми.

Последовательность наблюдавшихся значений, полученных повторением заданной операции, не будет обычно статистически контролируемой, даже если основные условия и предполагаются одинаковыми. Для того, чтобы узнать, являются ли основные условия одинаковыми, всегда нужно проверить, можно ли считать последовательность случайной, чтобы применить статистические методы.

Многие величины или факторы, вызывающие изменения в течении тех или иных биологических процессов, не могут быть полностью подконтрольными. Эти причины могут вызывать отклонения или внезапные скачкообразные изменения в функционировании систем организма, в результате чего изменения в состоянии здоровья нельзя предсказать.

По терминологии Walter A. Shewhart [1], система (неизвестных) причин, соответствующих статистически подконтрольному состоянию процесса, называется постоянной системой случайных причин, а неизвестные причины изменчивости, которые нарушают это состояние, называются привнесенными причинами. Опыт показывает, что с помощью статистических критериев можно обнаружить неподконтрольность процесса, найти привнесенные причины и сделать их подконтрольными так, чтобы процесс постепенно становился статистически подконтрольным.

Рассматривается система  $n$  различных наблюдавшихся значений  $x_1, x_2, \dots, x_n$  и последовательность знаков (“+“ или “-“) у  $(n-1)$  разностей  $x_{i+1} - x_i$ . Последовательность следующих друг за другом знаков “+” называется восходящей серией, а последовательность знаков “-” – нисходящей серией. Длина серии задается числом одинаковых знаков, определяющих ее. Общее число серий обозначается буквой  $R$ , число серий длины  $i$  – буквой  $r_i$ , и число серий длины, большей или равной  $k$ , – буквой  $R_k$ .

$$R_k = \sum_{i=k}^{n-1} r_i \quad (1)$$

Предполагая, что все  $n!$  возможных расположений из  $n$  чисел встречаются с одинаковой вероятностью, можно получить следующее среднее число серий:

$$M\{r_i\} = \frac{2}{(i+3)!} [n(i^2 + 3i + 1) - (i^3 + 3i^2 - i - 4)] \quad i \leq n-2 \quad (2)$$

где  $M$  – математическое ожидание,

$$\text{и } M\{R_k\} = \frac{2}{(k+2)!} [n(k+1) - (k^2 + k - 1)] \quad k \leq n-1 \quad (3)$$

Из табл. 1 можно найти число наблюдавшихся значений, дающее для данного  $k$  среднее  $M\{R_k\}=1$ . Для  $k=5$ , например,  $n=2549/6 \approx 425$ , то есть, в случайных расположениях из 425 наблюдений ожидается появление в каждом расположении только одной восходящей или нисходящей серии длины, большей или равной 5.

Для общего числа серий справедливы следующие выражения

$$M \{ R \} = \frac{1}{3} (2 n - 1) \tag{4}$$

$$V \{ R \} = \frac{1}{90} (16 n - 29) \tag{5}$$

$$V \{ R \} = \frac{n}{4} \left( 1 - \frac{1}{n - 1} \right) \tag{6}$$

где  $V(R)$  – дисперсия.

Для  $n > 20$  можно с хорошим приближением считать  $R$  распределенным нормально.

Критерий значимости, основанный на длинных сериях, может быть найден с помощью пуассонова приближения, так как для  $k \geq 5$

$$V \{ R_k \} \approx M \{ R_k \} \tag{7}$$

Это приводит к следующему приближению (для  $k \geq 5$ ):

$$P \{ R_k \geq 1 \} \approx 1 - e^{-M \{ R_k \}} \tag{8}$$

Подставляя значение  $M\{R_k\}$  из выражения (3) и решая относительно  $n$ , можно получить

$$n \approx k - \frac{1}{k + 1} - \frac{k + 2}{2} k ! \ln (1 - P \{ R_k \geq 1 \}) \tag{9}$$

Таким образом, для заданных  $P$  и  $k$  можно определить значение  $n$ , которое соответствует вероятности  $P$  того, что случайное расположение будет содержать, по крайней мере, одну серию длины, большей или равной  $k$ .

Из (9) следует:

$$P\{R_5 \geq 1\} \leq 0.05 \quad \text{для } n \leq 26 \tag{10}$$

$$P\{R_6 \geq 1\} \leq 0.05 \quad \text{для } n \leq 153 \tag{11}$$

$$P\{R_7 \geq 1\} \leq 0.05 \quad \text{для } n \leq 1170 \tag{12}$$

Значит, можно ожидать, что в выборках объема меньшего, чем 1170, серии длины, большей или равной 7, будут встречаться очень редко.

Гипотезой, альтернативной гипотезе о статистической подконтрольности, является гипотеза о постепенном смещении распределения в течение извлечения выборки. Такое изменение будет вызывать смещения или колебания в наблюдаемых значениях, так что можно ожидать появления одной или более длинных серий и что общее число серий будет небольшим. Следовательно, гипотеза о подконтрольности может быть проверена с помощью общего числа серий (значимым будет малое число серий) и, далее, с помощью длины серий (значимым будет появление очень длинных серий) [2].

Важнейшим свойством метода восходящих и нисходящих серий является независимость от характера распределения. Исходя из этого, его можно отнести к классу непараметрических.

Таблица 1

**Математическое ожидание числа восходящих и нисходящих серий длины, большей или равной  $k$ , в случайных расположениях из  $n$  различных чисел**

Длина серии, $k$	Математическое ожидание, $M\{R_k\}$
1	$1/3 \times (2n - 1)$
2	$1/12 \times (3n - 5)$
3	$1/60 \times (4n - 11)$
4	$1/360 \times (5n - 19)$
5	$1/2520 \times (6n - 29)$
6	$1/20160 \times (7n - 41)$
7	$1/181440 \times (8n - 55)$

С помощью метода восходящих и нисходящих серий можно обнаружить наличие привнесенных причин, узнать, в какой момент эти причины действуют и какого типа действие они производят. А это может помочь исследователю опознать привнесенные причины и подвести их под контроль.

При определении статистической подконтрольности восьмилетней динамики артериального давления в трех группах детей с нормо-, гипо- и гипертензией, проживающих в различных экологических условиях на территории Республики Беларусь было установлено следующее.

Для группы детей с гипотензией математическое ожидание и число наблюдаемых серий совпали (5 и 5), и даже при значительной дисперсии (1.1) достоверность приближения процесса к статистически подконтрольному высока –  $\approx 0.69$ . Длинных серий не наблюдается, что для расположения из 8 наблюдений также свидетельствует о подконтрольности.

Для групп с нормотензией и гипертензией число наблюдаемых серий отличается от математического ожидания (математическое ожидание – 5, число серий – 4) и при дисперсии 1.1 степень приближения процесса к подконтрольному  $\approx 0.32$ . Число длинных серий незначимо. Таким образом, на групповую динамику в двух последних случаях влияет некоторый привнесенный фактор (факторы). Здесь таким фактором может являться двусторонний обмен между группами детей с нормотензией и гипертензией с течением времени. Это нарушает однородность групп и, очевидно, пре-

вращает процесс при данном способе выборки в статистически неподконтрольный.

Еще один пример использования метода – проверка статистической подконтрольности индивидуальной динамики частоты сердечных сокращений по данным кардиоинтервалографии (КИГ) в различных физиологических условиях (лежа, в ортостазе и после физической нагрузки) (табл. 2).

Из приведенных в табл. 2 данных следует, что в значительной степени статистически подконтрольной является только динамика частоты сердечных сокращений в положении лежа. На это указывают следующие факты: число наблюдаемых серий в случае с пациентом А больше (73), а в случае с пациентом Б незначительно меньше (63) ожидаемого (65); коэффициенты значимости (0.97 и 0.3 соответственно) достаточно высоки; длинные серии отсутствуют. Динамика ЧСС в ортостазе в обоих случаях оказалась статистически неподконтрольной (число наблюдаемых серий – 76 и 90 – значительно меньше ожидаемого – 111 и 128 соответственно; коэффициенты значимости равны нулю; наблюдается большое количество длинных и очень длинных серий), и статистические методы в данном случае неприменимы. После физической нагрузки динамика ЧСС в данных примерах также остается статистически неподконтрольной (число наблюдаемых серий меньше ожидаемого, коэффициенты значимости равны 0.04 и 0 соответственно, и в случае с пациентом Б имеют место семь длинных серий) (табл. 2).

Таблица 2

**Результаты применения метода восходящих и нисходящих серий для проверки статистической подконтрольности индивидуальной динамики частоты сердечных сокращений (ЧСС)**

Вид регистрации КИГ	Статистические показатели				
	Объем выборки интервалов R-R	Математическое ожидание общего числа серий	Число наблюдаемых серий (из них длиной $\geq 4$ )	Дисперсия	Значимость подконтрольности
Пациент А лежа	99	65	73 (0)	17.28	0.97
	167	111	76 (5)	29.37	0.0
	после физической нагрузки	150	99	90 (1)	26.34
Пациент Б лежа	99	65	63 (0)	17.28	0.3
	193	128	90 (11)	33.99	0.0
	после физической нагрузки	170	113	90 (7)	29.90

Отсутствие статистической подконтрольности индивидуальной динамики ЧСС в ортостазе и после физической нагрузки указывает на наличие привнесенных серьезных факторов, значимо влияющих на процесс и требующих логического осмысления. Наиболее вероятным в ортостазе является воздействие гравитации в качестве внешнего естественно привнесенного фактора, выражено влияющего на уровень артериального давления и соответственно на ЧСС, как один из факторов его формирующих. В случае воздействия физической нагрузки добавляются экстракардиальные факторы кровообращения (венозная pompa, микронасосная функция скелетных мышц, присасывающее действие грудной клетки и др.), которые также являются привнесенными причинами, нарушающими статистическую подконтрольность процесса.

Таким образом, использование индивидуальной динамики ЧСС, получаемой в процессе регистрации КИГ, возможно только для оценки исходного вегетативного статуса, но не для определения вегетативной реактивности и вегетативного обеспечения деятельности.

Метод восходящих и нисходящих серий использовался для оценки статистической подконтрольности случайных выборок, получаемых в процессе эпидемиологических срезовых (cross-sectional) исследований. У обследованных детей определялась подконтрольность выборки величин экскреции йода с мочой. Для первой выборки число наблюдавшихся серий – 294 – превысило математическое ожидание, составляющее 278 серий и при дисперсии  $\approx 73.99$  степень приближения к подконтрольности высока ( $\approx 0.97$ ). Для второй выборки число серий (247) меньше ожидаемого (255) и при дисперсии  $\approx 67.94$  степень приближения к подконтрольности значительно меньше и приближенно составляет 0.17. Число длинных серий незначимо

в обеих выборках. Следовательно, во втором случае существуют не учтенные при формировании выборки причины, влияющие на ее подконтрольность. Главной причиной может быть некомпактное проживание детей и, как следствие, различное содержание стабильного йода в питьевой воде, потребляемой пище и различные рационы питания. Кроме того, причиной статистической неподконтрольности может быть неоднородность (в силу различных причин – генетических, демографических, физиологических и т.д.) выборки и скоррелированность показателей.

Аналогично были проверены на статистическую подконтрольность три выборки активности радионуклидов  $^{137}\text{Cs}$  в организме детей, проживающих в д. Ольманы, Лунинецком районе и Гомельской области (табл. 3).

Полученные результаты указывают на то, что все три выборки статистически неподконтрольны, о чем в каждой из них свидетельствуют число серий меньше ожидаемого и низкий коэффициент значимости. Кроме того, в выборке по Гомельской области присутствует серия длиной 7, вероятность появления которой в данном случае крайне мала. Это указывает на наличие выраженных привнесенных причин (факторов), которые должны быть определены и учтены в дальнейших исследованиях.

Таким образом, метод восходящих и нисходящих серий может эффективно использоваться при обработке широкого спектра медико-биологических и экологических данных [3–8].

Простота и отсутствие сложных математических вычислений делают метод доступным для использования в медицинских научно-практических исследованиях.

Первым этапом статистической обработки медико-биологических данных должна быть оценка их статистической подконтрольности.

Таблица 3

**Результаты применения метода восходящих и нисходящих серий для проверки статистической подконтрольности выборок показателей счетчика инкопорированных частиц (СИЧ)**

Район проживания	Статистические показатели				
	Объем выборки показателей СИЧ	Математическое ожидание общего числа серий	Число наблюдаемых серий	Дисперсия	Значимость подконтрольности
д. Ольманы	253	168	154	44.66	0.02
Лунинецкий район	218	145	140	38.43	0.23
Гомельская область	392	261	252	69.37	0.16

Метод восходящих и нисходящих серий должен быть использован для определения статистической подконтрольности изучаемых медико-биологических процессов и явлений.

Использование статистических критериев для обработки медико-биологических данных, в том числе метода восходящих и нисходящих серий, и оценка результатов их применения должны быть логически обоснованы.

Простота метода восходящих и нисходящих серий позволяет рекомендовать его к применению для экспресс-оценки и формирования предварительного прогноза развития событий в условиях чрезвычайных ситуаций различного характера, обусловленных широким спектром факторов – радиационных, химических, экологических, эпидемиологических, социологических и т.д. Данный метод может быть эффективно использован при выявлении таких ситуаций на ранней – скрытой – стадии их развития.

#### Литература

1. *Shewhart, W.A.* Economic Control of Quality of Manufactured Product / W.A.Shewhart. – New York, 1931. – Appendix II.
2. *Хальд, А.* Математическая статистика с техническими приложениями / А.Хальд. – М.: Иностранная литература, 1956. – 664 с.
3. *Бубнов, В.П.* Решение задач экологического менеджмента с использованием методологии системного анализа / В.П.Бубнов, С.В.Дорожко, С.А.Лаптёнок. – Минск: БНТУ, 2009. – 266 с.
4. *Лаптёнок, С.А.* Оценка статистической подконтрольности медико-биологических процессов / С.А.Лаптёнок, А.Н.Аринчин, Н.В.Арсюткин // *Здравоохранение*. – 1998. – №1. – С.29–31.
5. *Лаптёнок, С.А.* Применение метода восходящих и нисходящих серий для анализа статистической подконтрольности медико-биологических данных. Методические рекомендации (утв. МЗ РБ 23.12.1998) / С.А.Лаптёнок, А.Н.Аринчин, Н.В.Арсюткин. – Минск: МЗ РБ, 1999. – 31 с.
6. *Лаптёнок, С.А.* Информационно-аналитический комплекс для математической обработки медико-экологических данных в целях решения задач по минимизации последствий чрезвычайных ситуаций: автореф. дис. ... канд. техн. наук: 05.26.02 / С.А.Лаптёнок; ИРБ «БЕЛРАД». – Минск, 2001. – 23 с.
7. *Лаптёнок, С.А.* Системный анализ геоэкологических данных в целях митигации чрезвычайных ситуаций. – Минск: БНТУ, 2013. – 287 с.
8. *Моделирование экологических процессов. Учебное пособие / С.А.Лаптёнок, И.В.Лазар, Н.М.Новикова [и др.].* – Минск: ИВЦ Минфина, 2022. – 144 с.

#### STATISTICAL CONTROLLABILITY ASSESSMENT IN MEDICAL AND ECOLOGICAL RESEARCH

<sup>1</sup>S.A.Laptyonok, <sup>1</sup>O.I.Rodzkin,  
<sup>2</sup>A.A.Kologrivko, <sup>2</sup>S.A.Horeva, <sup>2</sup>Yu.V.Kliausava

<sup>1</sup>International Sakharov Environmental Institute of Belarusian State University, 23/1, Dolgobrodskaya Str., 220070, Minsk, Republic of Belarus

<sup>2</sup>Belarusian National Technical University, 65, Nezavisimosti Ave., 220013, Minsk, Republic of Belarus

Effectiveness evaluation results on data statistical controllability assessment in medical and ecological research using method of ascending and descending series are presented in the article.

Keywords: statistical controllability; method of ascending and descending series; medical and ecological research.

#### Сведения об авторах:

**Лаптёнок Сергей Антонович**, канд. техн. наук, доцент; Международный государственный экологический институт имени А.Д.Сахарова Белорусского государственного университета, факультет мониторинга окружающей среды, доцент кафедры ядерных и медицинских технологий; тел.: (+37529) 6257828; e-mail: 267413@mail.ru.

**Родькин Олег Иванович**, д-р биол. наук, доцент; Международный государственный экологический институт имени А.Д.Сахарова Белорусского государственного университета, директор; тел.: (+37529) 3871501; e-mail: aleh.rodzkin@rambler.ru.

**Кологривко Андрей Андреевич**, канд. техн. наук, доцент; Белорусский национальный технический университет, факультет горного дела и инженерной экологии, декан; тел.: (+37529) 6840510; e-mail: akologrivko@bntu.by.

**Хорева Светлана Алексеевна**, д-р биол. наук, профессор; Белорусский национальный технический университет, факультет горного дела и инженерной экологии, профессор кафедры «Инженерная экология»; тел.: (+37529) 6857939; e-mail: horeva.svetlana@mail.ru.

**Кляусова Юлия Владимировна**, канд. с.-х. наук, доцент; Белорусский национальный технический университет, факультет горного дела и инженерной экологии, доцент кафедры «Инженерная экология»; тел.: (+37529) 1679237; e-mail: yuliya-klaus@mail.ru.

Поступила 16.05.2024 г.