

УДК 574: (- 056.32)

**МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ: ПЕРВЫЕ ШАГИ
АПРОБАЦИИ ДЛЯ РЕШЕНИЯ МЕДИКО-ЭКОЛОГИЧЕСКИХ ПРОБЛЕМ НА
ТЕРРИТОРИИ КРЫМА**

Евстафьева Е.В., Овсянникова Н.М.

*Крымский государственный медицинский университет им. С.И. Георгиевского, Симферополь,
Украина
E-mail: nataly@csmu.strace.net*

Описаны принципы использования методов многомерного статистического анализа в количественной оценке влияния факторов окружающей среды на здоровье человека. Оценены возможности и перспективы использования кластерного, факторного и множественного регрессионного анализов для решения задач экологической физиологии, а также нейросетевого подхода в описании зависимости «доза – время – эффект».

Ключевые слова: многомерный статистический анализ, факторы окружающей среды, здоровье человека.

В последние годы медицина и биология вступили в новую фазу своего развития. Накопление огромных массивов количественных данных и доступность вычислительной техники усилило математизацию этих научных сфер. На базе накопленных фактов исследуются количественные закономерности биологических процессов, создаются математические модели исследуемых явлений и объектов, а применение математической статистики в работах медико-биологического направления даёт возможность решить вопрос о мере достоверности получаемых результатов [1]. При этом роль используемых статистических методов двояка. С одной стороны они помогают обнаруживать ранее неизвестные закономерности, с другой стороны – с их помощью авторы проверяют достоверность априорно формулируемых гипотез. Однако некорректность применения статистики, использование авторами методов, не адекватных задачам исследования даже на одном из этапов работы, делает весьма сомнительными, а иногда просто несостоятельными выводы [2]. В связи с этим применение математического аппарата в решении медико-биологических задач, значительно расширяя возможности исследователей в оценке и интерпретации получаемых результатов, требует корректного их использования и соответствующей математической «грамотности» медиков и биологов, в особенности, когда речь идет о наиболее сложных областях научно-практического знания.

К числу таких научных направлений относится экология человека, в особенности тот ее раздел, где производится количественная оценка влияния факторов окружающей среды на здоровье человеческой популяции. Для решения

этой задачи необходимо использование многомерных методов вариационной статистики, с помощью которых будет производиться не только сложная сама по себе количественная оценка экологической ситуации, количественная оценка здоровья человеческой популяции, но оценка взаимодействия этих многомерных массивов данных.

Уровень сложности решаемых в этой области задач может существенно различаться и требовать достаточно высокой компетентности исследователя не только в области математической статистики, но и математического моделирования [3], что часто затрудняет, а иногда делает невозможной корректную оценку влияния окружающей среды на здоровье населения и ограничивает число реально решаемых задач [4].

Среди них выявление экологической обусловленности заболеваний представляет в последнее время исключительный интерес в связи с увеличением экологического риска для здоровья и необходимостью принятия действенных мер по его снижению. Оценка здоровья населения в связи с экологической ситуацией возможна при проведении натуральных (мониторинговых) исследований различного уровня: глобального, регионального, субрегионального, локального. Каждый из них позволяет решать различные задачи: региональный мониторинг здоровья и экологической ситуации дает возможность интегральной оценки здоровья на крупномасштабных территориях; субрегиональный – позволяет оценить специфику состояния здоровья на отдельных территориях, определить приоритетные по влиянию на здоровье факторы окружающей среды. Именно в результате таких исследований возможна объективная количественная оценка техногенного воздействия на организм [5].

Количественная интерпретация результатов мониторинга должна включать: общую характеристику пространственно-временной изменчивости состояния окружающей среды и здоровья; выявление приоритетных по степени негативного воздействия на здоровье и распространенности в окружающей среде экологических факторов и оценку силы и характера их влияния на организм; прогнозирование изменений в состоянии здоровья населения в связи с изменением экологической ситуации [5-7].

В свое время методологические основы и методические подходы к такого рода мониторинговым исследованиям были разработаны и изложены в [6]. Изложение результатов внедрения некоторых из них с использованием соответствующего математического аппарата в крымском регионе и явилось целью настоящей статьи.

Региональные мониторинговые исследования в настоящее время проводятся в следующих направлениях: количественная оценка экологической ситуации; оценка состояния здоровья населения в целом по Крыму и на отдельных территориях; оценка влияния факторов среды на здоровье выборочных контингентов городского и сельского населения.

Для количественной оценки экологической ситуации освоены и апробированы рекомендованные Европейской конвенцией о трансграничных переносах атмосферных загрязнителей методики расчета допустимых критических нагрузок и их превышений загрязнителями (тяжелыми металлами, окислами азота и серы,

пестицидами) с учетом природных особенностей региона и влияния на здоровье человека [8]. С использованием геоинформационных технологий произведено картографирование критических нагрузок тяжелых металлов (свинец, кадмий, ртуть) для лесных и сельскохозяйственных территорий АР Крым для координатной сетки 5x5 км. Пример такого картографирования приведен на Рис. 1. Совместно с Республиканским Комитетом по охране окружающей природной среды произведено определение фактических выпадений тяжелых металлов на территории южного региона Крыма (рис. 2) и оценка превышений допустимых нагрузок (рис. 3).



Рис. 1. Критические нагрузки кадмия на экосистемы АР Крым.

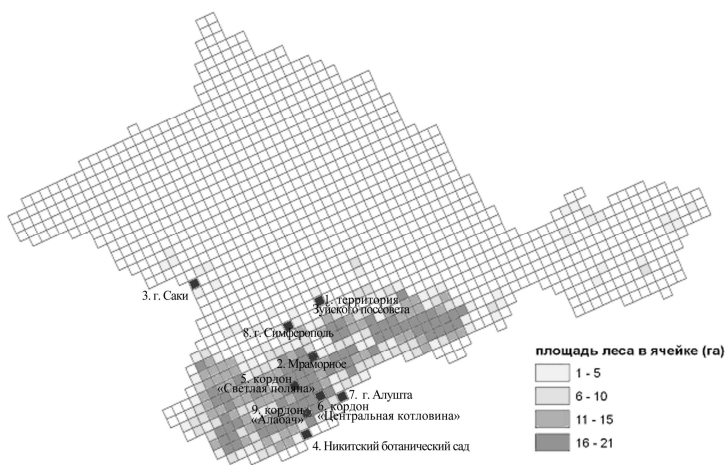


Рис. 2. Расположение экспериментальных площадок для определения превышений критических нагрузок в южном регионе АР Крым.

Для интегральной количественной оценки состояния здоровья населения Автономной республики Крым проанализированы временные и пространственные распределения заболеваемости и выявлены территории первоочередной оценки риска для здоровья [9]. Примеры такого временного и пространственного анализа приведены на Рис. 4-5.

Таким образом, мониторинговые исследования такого регионального и субрегионального уровня позволили в целом дать интегральную характеристику состояния здоровья на территории Автономной республики Крым и оценить экологическую ситуацию пока только в основном в отношении наиболее опасных тяжелых металлов и не для всей территории Крыма.

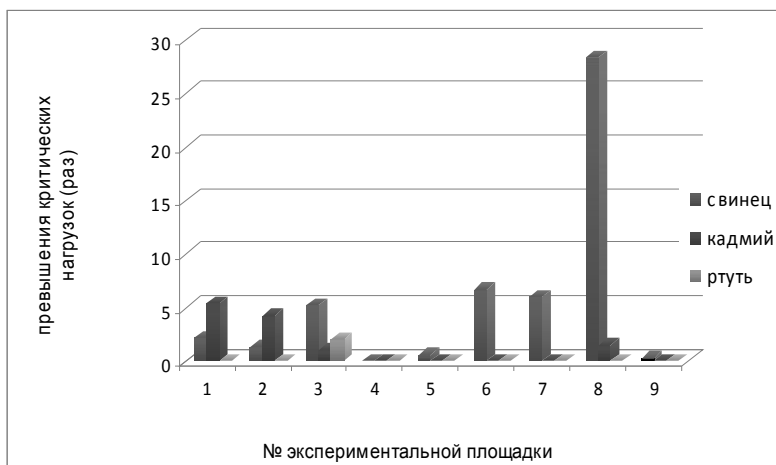


Рис. 3. Превышения критических нагрузок на экспериментальных площадках.



Рис. 4. Временной анализ заболеваемости на территории АР Крым

Динамика болезненности в 90-х годах

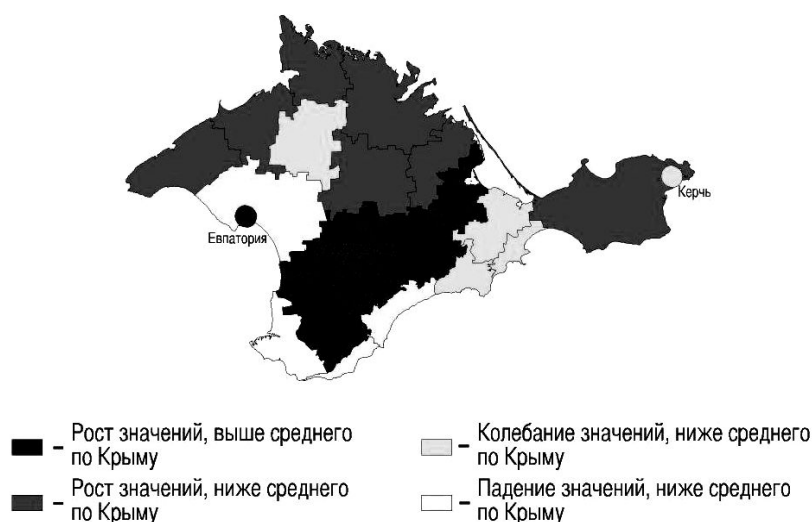


Рис. 5. Пространственный анализ заболеваемости на территории АР Крым.

Оценка влияния экологической ситуации на здоровье требует использования многомерного статистического анализа [10]. На сегодняшний день в исследовании многомерных признаков пространств выделяют три основных направления: задачи многомерной классификации, сжатое представление данных и статистическое исследование зависимостей [10].

Разнообразные задачи многомерной классификации, или разделение выборочной совокупности объектов в многомерном пространстве, осуществляются методами кластерного или дискриминантного анализа. Их решение позволяет разделить весь массив данных на отличающиеся между собой, но однородные внутри, группы, основываясь на значениях внутри- и межгрупповых дисперсий.

Кластерный анализ – это совокупность методов, позволяющих классифицировать многомерные наблюдения, каждое из которых описывается набором исходных переменных. Целью кластерного анализа является образование групп схожих между собой объектов, которые принято называть кластерами.

Необходимость развития методов кластерного анализа и их использования продиктована прежде всего тем, что они помогают построить научно обоснованные классификации, выявить внутренние связи между единицами наблюдаемой совокупности. Кроме того, методы кластерного анализа могут использоваться с целью сжатия информации, что является важным фактором в условиях постоянного увеличения и усложнения потоков статистических данных.

Преимущество кластеризации в решении задач экологической физиологии заключается в том, что разделение на группы и последующее сравнение осуществляется не по какому-либо одному показателю, а по всему их набору [7].

Метод кластерного анализа был применен при анализе данных медицинской статистики по заболеваемости населения Автономной республики Крым за период 1993-2008 гг.

Для выполнения кластеризации в кластерном анализе используются следующие методы: Joining (tree clustering) (древовидная кластеризация), K – means clustering (метод K средних), Two-way joining (двухходовое объединение).

Методом древовидной кластеризации районы Крыма разделились на три кластера, достоверно отличающихся друг от друга по структуре и уровню заболеваемости (высокая, средняя и низкая степени) (рис. 6). Интересно отметить, что в первый кластер с высоким уровнем заболеваемости и одинаковым нозологическим профилем вошли Симферополь, Евпатория и Ялта.

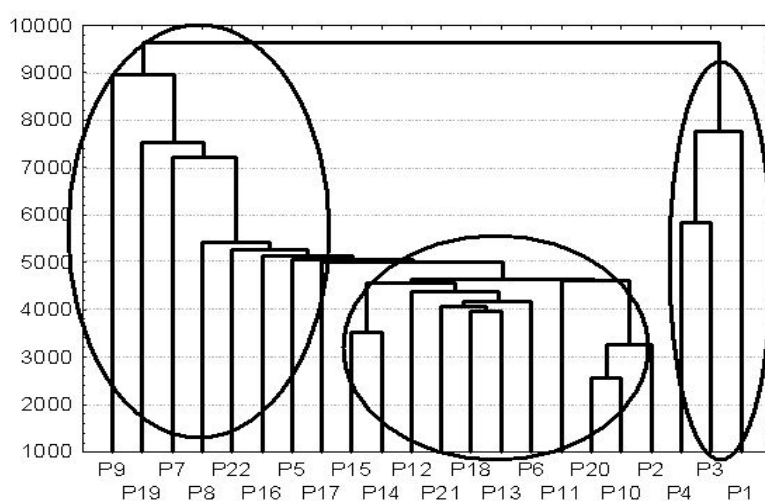


Рис. 6. Древовидная кластеризация территории АР Крым по заболеваемости в регионах (P1–P21).

Второе направление в исследовании многомерных признаков пространств, – сжатое представление данных, – используется для уменьшения объема информации в многомерных моделях как по числу объектов, так и по числу признаков и выделения подмножества наиболее значимых признаков из их исходного множества. Это предположение исходит из гипотезы о наличии в них определенных внутренних закономерностей, в качестве внешних проявлений которых выступают наблюдаемые признаки. С этой целью используют факторный анализ (ФА).

В современной статистике под факторным анализом понимают совокупность методов, которые на основе реально существующих связей признаков (или

объектов) позволяют выявлять латентные обобщающие характеристики организационной структуры и механизма развития изучаемых явлений и процессов.

ФА последовательно решает две взаимосвязанные задачи: 1) сжимает исходный массив эмпирических данных, выражая их в терминах относительно небольшого числа независимых переменных (факторов); 2) вскрывает функциональные зависимости между исходными признаками, описывая сходство и различие в терминах выделенных факторов [11].

В задаче мониторинга ФА может быть использован именно как метод отбора переменных, наиболее адекватно отображающих ту информацию, которая содержится во всем наборе близких по характеру групп показателей [7].

Техника ФА направлена на оценку факторных нагрузок и специфических дисперсий. Чем выше абсолютное значение факторной нагрузки, тем теснее связь соответствующей переменной с данным фактором, тем больше ее вес в его структуре. Тем самым совокупность факторных нагрузок конкретной переменной по всем выделенным факторам свидетельствует о том, какая доля вариации данной переменной вызвана каждым фактором [11].

Для выявления факторов, наиболее существенно влияющих на заболеваемость населения АР Крым, весь массив данных медицинской статистики был подвергнут факторному анализу. Он был произведен для каждого кластера в отдельности и позволил определить количество гипотетических факторов, значимо влияющих на структуру заболеваемости в каждом кластере и факторные нагрузки. Для примера приведены критерий отбора числа значимых факторов (рис. 7) для 1-го кластера. Как следует из рисунка, уровень заболеваемости в первом кластере обусловлен двумя факторами. При этом один из них, судя по его величине, играет определяющую роль.

Факторный и кластерный анализы были использованы и другими авторами в задачах количественной оценки экологического риска: многофакторный анализ канцерогенной опасности радона [12], комплексная оценка состояния здоровья населения и контроля качества объектов окружающей среды [13, 14], анализе влияния экологических факторов на здоровье населения [15-18].

Третье направление – статистическое исследование зависимостей используется для исследования характера и структуры взаимосвязей, существующих между анализируемыми показателями. К этому направлению относятся методы регрессионного, корреляционного, дисперсионного и ковариационного анализа, которые используют в настоящее время чаще всего для решения задач на локальном уровне мониторинговых исследований.

В задачах экологической физиологии и экологического нормирования, решаемых на этом уровне, оценка влияния отдельных загрязнителей на организм человека часто производится по величине простых корреляционных связей. Так, на территории АР Крым с использованием корреляционного анализа произведено исследование влияния тяжелых металлов на функционирование нервной, вегетативной, иммунной и сердечно-сосудистой систем детей и взрослых в г. Симферополе и на промышленно загрязненной территории Северного Крыма [19,

20], который позволил обнаружить нейро-, вегето-, кардио- и иммунотропные эффекты отдельных металлов.

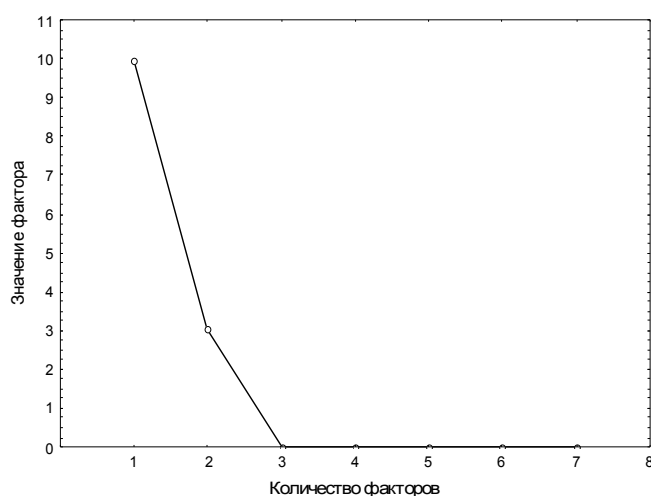


Рис. 7. Количество значимых факторов для 1-го кластера.

В то же время в реальных условиях организм подвергается одновременному воздействию множества факторов среды и каждый из показателей здоровья обычно связан не с каким-либо одним из загрязнителей, а с их «набором». Известно, что их комплексное действие может иметь синергический, антагонистический и аддитивный характер [6]. В этой связи возникает проблема выбора такого подмножества загрязнителей, которые наиболее значимо влияют на тот или иной показатель. Для решения такого рода задач используют множественный регрессионный анализ (МРА). Оптимальным набором регрессоров следует считать тот, который обеспечивает объяснение наиболее существенной доли дисперсии изучаемого показателя [7].

В рамках эколого-физиологического мониторинга, проводимого в АР Крым, посредством МРА оценено действие комплекса токсичных и эссенциальных металлов на состояние иммунитета [21].

Результаты анализа позволили выявить достоверное влияние комплекса указанных металлов на отдельные показатели и, в целом, состояние иммунной системы детей и взрослых. Суммарный эффект влияния металлов был более весомым и существенно отличался от их индивидуального действия, о чем свидетельствовала большая величина R-коэффициентов множественной регрессии по сравнению с β -коэффициентами регрессии, которые отражают вклад отдельного металла в комплексное действие (табл. 1). Кроме того, направленность влияния (прямые или обратные связи) в ряде случаев свидетельствовала об антагонистических отношениях между токсичными и эссенциальными металлами, а в некоторых случаях – о потенцировании совместного действия.

Таблица 1
Данные множественного регрессионного анализа показателей иммунной системы и содержания металлов в волосах детей в возрасте 1-6 лет

Показатели иммунной системы	Коэффициент множественной корреляции, R	Коэффициенты регрессии для отдельных металлов (β -коэффициенты)				
		Pb	Cd	Zn	Cu	Hg
О-лимф. (относ. кол.)	0,609 p<0,09				-0,595 p<0,01	
CD3	0,619 p<0,07					0,644 p<0,02
CD3 (относ. кол.)	0,785 p<0,002				0,734 p<0,0003	0,542 p<0,01
CD4	0,635 p<0,06					0,674 p<0,01
CD4 (относ. кол.)	0,631 p<0,06	0,545 p<0,02				
CD8	0,612 p<0,08					0,571 p<0,03
CD8 (относ. кол.)	0,691 p<0,02		-0,581 p<0,03		0,814 p<0,001	
Лейкоциты	0,792 p<0,001					0,787 p<0,001
IgM	0,581 p<0,09	0,526 p<0,03				
Палочко-ядерные нейтрофилы	0,628 p<0,07		0,724 p<0,01			

По результатам МРА были составлены уравнения множественной регрессии для выявленных достоверных коэффициентов множественной регрессии. Такие математические модели могут иметь практический интерес в решении задач прогнозирования состояния здоровья при воздействии негативных факторов окружающей среды.

Метод МРА широко использовался и другими авторами в эколого-физиологических и гигиенических исследованиях: в гигиенической оценке формирования суммарного риска популяционному здоровью на урбанизированных территориях [22], в комплексной оценке состояния здоровья населения [13], в экологическом мониторинге оценки и контроля качества объектов окружающей среды [14, 23], в оценке комплексного влияния токсичных металлов на иммунную систему детей и взрослых, проживающих на урбанизированной и промышленно загрязненной территориях [24-26].

ВЫВОД

Таким образом, апробация выше изложенных подходов количественной оценки влияния факторов окружающей среды на здоровье человека на территории Автономной республики Крым позволила получить некоторые предварительные результаты о влиянии факторов окружающей среды и показала перспективность их дальнейшего использования в последовательном решении задач мониторинговых исследований различного уровня. Перспективным направлением в развитии применения методов многомерного анализа и их привлечение к решению региональных медико-экологических задач может явиться нейросетевой подход [27] и динамическое моделирование [2].

Список литературы

1. Кобринский Б.А. Принципы математико-статистического анализа данных медико-биологических исследований / Б.А. Кобринский // Росс. Вестник перинатологии и педиатрии – 1996. – №4. – С. 60–64.
2. Леонов В.П. Применение статистики в медицине и биологии: анализ публикаций 1990-1997 гг. Деп. ВИНТИ 23.01.98Х9 179-В98 / В.П. Леонов, П.В. Ижевский – ГНЦ РФ Институт биофизики М. – 1998. – 37 с.
3. Hettelingh J-P. Progress in the modelling of critical thresholds, impacts to plant species diversity and ecosystem services in Europe / Hettelingh J-P., Posch M., Slootweg J. – LRTAP. – 2009 – 128 p.
4. Антомонов М.Ю. Математичне забезпечення гігієнічних досліджень / М.Ю. Антомонов // Довкілля та здоров'я. – 2001. – №2. – С.57–58.
5. Евстафьева Е.В. Методологические основы и методологические подходы к экологическому нормированию техногенного воздействия на организм человека / Е.В. Евстафьева // Таврический медико-биологический вестник. – 2001. – Т. 4, № 4. – С. 7–11.
6. Евстафьева Е.В. Физиологическое и биохимическое обоснование проблемы адаптации человека к различным условиям среды обитания : автореф. дис. ...д.б.н.. / Е.В. Евстафьева– М.: УДН, 1996. – 32 с.
7. Опополь Н. Эколого-гигиенический мониторинг: проблемы и решения / Н. Опополь, Р. Коробов – Кишинев, 2001. – 240 с.
8. Hettelingh JP. Heavy Metal Emissions, Depositions, Critical Loads and Exceedances in Europe / JP.Hettelingh, J.Sliggers (eds) – LRTAP. – 2005 – 94 p.
9. Использование геоинформационных технологий в анализе пространственной неоднородности заболеваемости / Е.В. Евстафьева, А.А. Артов, М.Г. Щеголева [и др.] // Таврический медико-биологический вестник. – 2001. – Т. 4, № 4. – С. 91–96.
10. Айвазян С.А. Прикладная статистика: Исследование зависимостей: Справочник / Айвазян С.А., Енюков И.С., Мешалкин Л.Д. – М.: Финансы и статистика, 1985. – 182с.
11. Репин С.В. Математические методы обработки статистической информации с помощью ЭВМ. / С.В. Репин, С.А. Шейн – Минск : Университетское – 1990. – 128 с.
12. Лежин В.Л. Метод многофакторного анализа канцерогенной опасности радона / В.Л. Лежин, Е.В. Ползик, В.С. Казанцев // Гигиена и санитария. – 2008. – №1. – С. 79–82.
13. Мусийчук Ю.И. Опыт комплексной оценки состояния здоровья населения Санкт-Петербурга / Ю.И. Мусийчук, О.П. Ломов, В.М. Кудрявцев // Гигиена и санитария. – 2008. – №2. – С.15–17.
14. Руш Е.А. Экологический мониторинг : методы многомерного статистического анализа оценки и контроля качества объектов окружающей среды Приангарья / Е.А. Руш, А.Ю. Давыденко // Инженерная экология. – 2006. – № 5. – С. 3–26.
15. Полоусова Г.Ю. Статистический анализ влияния экологических факторов на здоровье населения Тульской области : автореф. дис. ... к.э.н. / Г.Ю. Полоусова – 2003. – 20 с.
16. Evaluation of xenobiotic impact on urban receiving waters by means of statistical methods / A. Musolff, S. Leschik, MT. Schafmeister [et al.] // Water Sci Technol. – 2010. – Vol. 62(3) – P. 684–692.

17. Characterization of trace metals of risk to human health in airborne particulate matter (PM2.5) at two sites in Guadalajara, Mexico / H. Saldarriaga-Nore, L. Hernandez-Mena, M. Ramirez-Mu [et al.] // *Environ Monit.* – 2009. – Apr;11(4) – P. 887–894.
18. Some potential hazardous trace elements contamination and their ecological risk in sediments of western Chaohu Lake, China / LG. Zheng, GJ. Liu, Y. Kang [et al.] // *Environ Monit Assess.* – 2010. – Jul;166(1-4), P. 379–386.
19. Євстаф'єва І.А. Особливості функціонального стану центральної нервової та серцево-судинної систем в зв'язі з вмістом важких металів в організмі підлітків : дисс...к.б.н. / І.А. Євстаф'єва – Симферополь, 2003. – 130 с.
20. Слюсаренко А.Е. Імунологічна реактивність організму в різних умовах техногенного забруднення середовища важкими металами : дисс...к.б.н. / А.Е. Слюсаренко – Симферополь, 2003. – 157 с.
21. Овсянникова Н.М. Особливості адаптаційних реакцій людини в зв'язі з вмістом важких металів в організмі / Н.М. Овсянникова // *Ученые записки Таврического национального университета им. В.И.Вернадского.* – 2010. – Т.23 (62). – № 2. – С. 142–151.
22. Боев В.М. Гігієнічна оцінка формування сумарного ризику для здоров'я населення на урбанізованих територіях / В.М. Боев, В.Н. Дунаєв, Р.М. Шагєєв // *Гігієна і санітарія.* – 2007. – №5. – С. 79–82.
23. Сидоров П.І. Ментальна екологія: від концепцій залежних розладів до системного моніторингу здоров'я / П.І. Сидоров // *Медицина праці та промислова екологія* – 2007. – №2. – С. 1–9.
24. Lead induced DNA strand breaks in lymphocytes of exposed workers: role of reactive oxygen species and protein kinase C. / M.E. Fracasso, L. Perbellini, S. Sold [et al.] // *Mutat Res.* – 2002. – Mar; 515(1-2) – P. 159–169.
25. Muhammad S. Arsenic health risk assessment in drinking water and source apportionment using multivariate statistical techniques in Kohistan region, northern Pakistan / S. Muhammad, M. Tahir Shah, S. Khan // *Food Chem Toxicol.* – 2010. – Oct;48(10) – P. 2855–2864.
26. Sarasua S.M. Serum immunoglobulins and lymphocyte subset distributions in children adults living in communities assessed for lead and cadmium exposure / S.M. Sarasua, R.F. Vogt, L.O. Henderson // *Pediat. Allergy Immunol.* – 2000. – Vol. 8. – P. 161–167.
27. Сердюк А.М. Фактори середовища як фактори ризику для здоров'я населення: результати епідеміологічних досліджень, шляхи їх розширення та углиблення / А.М. Сердюк, Я.І. Звінячковський, О.В. Бердник // *Врачебное дело.* – 1996. – № 1–2. – С. 3–6.

Євстаф'єва О.В. Багатомірний статистичний аналіз: перші кроки апробації для вирішення медично-біологічних проблем на території Криму / О.В. Євстаф'єва, Н.М. Овсянникова // *Вчені записки Таврійського національного університету ім. В.І. Вернадського. Серія „Біологія, хімія”.* – 2010. – Т. 23 (62), № 4. – С. 78-88.

Описані принципи використання методів багатомірного статистичного аналізу в кількісній оцінці впливу чинників довкілля на здоров'я людини. Оцінено можливості та перспективи використання кластерного, факторного та множинного регресійного аналізів для вирішення завдань екологічної фізіології, а також нейронного підходу в описі залежності «доза - час - ефект».

Ключові слова: багатомірний статистичний аналіз, фактори навколишнього середовища, здоров'я людини.

Evstaf'eva N. Multivariate statistical analysis: a theoretical overview and first steps of approbation in the Crimea / N. Evstaf'eva, N. Ovsyannikova // *Scientific Notes of Taurida V.I. Vernadsky National University.* – Series: Biology, chemistry. – 2010. – Vol. 23 (62), No 4. – P. 78-88.

Individual psychological determinants of the personality types determined using the Jenkins' questionnaire are described. The coronary-prone type is psychologically characterized basing on the results of correlation and dispersion analyzes. Gender differences between the subjects of the coronary-prone type are shown.

Keywords: type A behavior, coronary-prone personality type, questionnaire by Jenkins.

Поступила в редакцію 18.11.2010 г.