

ПОДХОДЫ К СТАТИСТИЧЕСКОМУ МОДЕЛИРОВАНИЮ МЕДИКО-БИОЛОГИЧЕСКИХ ДАННЫХ НА ПРИМЕРЕ ОЦЕНКИ ФАКТИЧЕСКОГО ПИТАНИЯ ПОДРОСТКОВ

Е. Д. Константинова¹, Т. А. Маслакова¹, А. Н. Варакин¹, С. С. Делец²

¹Институт промышленной экологии УрО РАН, г. Екатеринбург, Россия

²Федеральное государственное бюджетное образовательное учреждение высшего образования «Южно-Уральский государственный медицинский университет» Министерства здравоохранения Российской Федерации (ФГБОУ ВО ЮУГМУ Минздрава России), г. Челябинск, Россия

В методической статье обсуждаются вопросы учета особенностей работы с медико-биологическими данными (МБД) при решении эпидемиологических задач с помощью статистических методов. Подчеркивается важность правильного выбора статистических методов анализа МБД, а также наглядного графического представления экспериментальных численных результатов из области эпидемиологии. Был ретроспективно проанализирован массив МБД, содержащий 809 случаев и 185 переменных разного типа. В качестве метода анализа использована множественная линейная регрессия с применением подхода, основанного на скользящем среднем. Решена практическая задача построения моделей, описывающих сопряженность индекса массы тела (ИМТ) подростков с пищевыми и некоторыми непищевыми факторами. Исследована распространенность дефицита массы тела, избыточной массы тела и ожирения среди подростков – учащихся колледжей. Результаты настоящего исследования могут быть востребованы при разработке профилактических мер по оптимизации питания в организованных коллективах, а также способствуют повышению уровня гигиенической грамотности заинтересованного круга лиц.

Ключевые слова: статистическое моделирование; медико-биологические данные; множественная линейная регрессия; фактическое питание подростков.

1. Введение

Развитие и массовое распространение прикладных статистических пакетов делает анализ МБД доступным практически для любого исследователя. Однако кажущаяся простота подобного анализа легко может ввести в заблуждение. Корректный результат может быть получен исключительно при правильном выборе статистических методов и учете особенностей работы с МБД. Обязательное соблюдение этих условий позволит исключить искажение и неправильное представление результатов.

Тема настоящей статьи – решение практической задачи эпидемиологии с помощью комплекса статистических методов. В предметном плане это означает поиск алгоритмов, которые будут выдавать ответ для каждого нового случая.

Стандартной задачей эпидемиологии является, например, следующее: есть список переменных, представляющих собой МБД, имеющих отношение к какой-либо медицинской проблеме, и требуется объяснить или спрогнозировать изменение одной из них. Для этого в биостатистике применяется статистическая модель – формальное представление взаимоотношений между переменными, которую потом можно использовать для желаемого объяснения или прогноза. Безусловно, на этом этапе должны появиться определенные математические идеи, но база должна быть понятна для специалиста предметной области (не математика). Другими словами, желательно получение интуитивно понятных решающих правил модели для использования специалистами – медиками и биологами.

МБД имеют ряд особенностей, которые необходимо учитывать при выборе

метода статистического моделирования. Достаточно подробно мы останавливались на них в предыдущем исследовании [1]. Так, вследствие зашумленности реальных МБД первоочередная задача при работе с ними – проверка на наличие «выбросов» – значений, не характерных для данной выборки. Все выбросы должны быть удалены для исключения искажения результатов анализа. В одной задаче анализа МБД исследователю зачастую приходится иметь дело не только с переменными разных типов (количественными, ранговыми и пр.), но и учитывать их коррелированность. При выборе метода, помимо типа переменных, необходимо принимать во внимание вид их распределения [2, 3].

Нельзя также обойти вниманием дизайн исследования. Как уже было сказано, основная цель в исследовании (при любом виде дизайна) – установление связи между предполагаемыми факторами риска (прогностическими факторами) и событиями (заболевание, смертность и т. п.). При этом вопрос о направленности причинно-следственной связи весьма сложен. В таком типе дизайна, как, например, одномоментное исследование, он спорен, поскольку фиксация определенного состояния и уровня некоторого фактора не позволяет судить о том, явился ли фактор причиной некоторого состояния исследуемого, или состояние вызвало определенный уровень исследуемого фактора [4].

Стоит отметить также, что любой тип дизайна сопровождается так называемыми случайными и систематическими ошибками – смещениями от истинных значений в силу некоторых особенностей исследования [5, 6].

Очевидна также невозможность получения 100 %-й точности результатов моделирования, поскольку мы имеем дело с «паутиной причинности» возникновения патологии. Ни один фактор риска возникновения болезни сам по себе не является непосредственной причиной заболевания, он может лишь увеличить или изменить вероятность его появления.

Что касается непосредственно выбора статистического метода, то про нежелательность применения моделей типа «черного ящика» с непрозрачной работой алгоритма говорит все большее число исследований [7, 8]. Получая решающее правило модели «черного ящика», мы не знаем, откуда взялось такое заключение и насколько мы можем быть в нем уверены: нам приходится либо принимать, либо отказываться от него. Не понимая, как алгоритм выдает ответ, мы не можем исследовать его на наличие неявных, но систематических ошибок.

Осознание необходимости в контролируемости алгоритмов, потенциально влияющих на жизнь людей, растет, и требования, чтобы выводы имели понятное объяснение, в некоторых странах уже включаются в законодательство. Так, в [9] одной из основных рекомендаций при выборе метода оценки профессионального риска названо обеспечение возможности прослеживания, воспроизводимости и проверки процесса и результатов, а также простоты и понятности применяемого метода.

Такие требования препятствуют использованию сложных закрытых алгоритмов при обработке МБД и могут приводить к предпочтению (довольно старомодных) регрессионных алгоритмов, которые обеспечивают математическое представление отношений между набором независимых (объясняющих) переменных и зависимой переменной (переменной отклика) [10–14].

На современном этапе одной из серьезных проблем человечества становится избыточная масса тела и ожирение, основа формирования которых – дисбаланс между потреблением веществ и реальными потребностями в них [15, 16]. Дети и подростки с отклонениями в состоянии питания составляют группу риска развития заболеваний пищеварительной, сердечно-сосудистой и костно-мышечной систем, а также сниженной резистентности [17–21]. Учащиеся профессионально-технических колледжей представляют собой особо уязвимую с точки зрения здорового образа жизни группу. Вместе с тем качество питания учащихся организаций среднего профессионального образования – фактор формирования будущего трудового потенциала общества.

В данной работе проведен анализ массива МБД с использованием методов статистического моделирования на примере данных о распространенности дефицита массы тела, избыточной массы тела и ожирения, а также сопутствующих пищевых и непищевых факторов среди подростков – учащихся колледжей. Результаты представлены в наглядной форме, доступной для специалистов – не математиков.

2. Материалы и методы

Был ретроспективно проанализирован массив МБД, содержащий 809 случаев и 185 переменных разного типа. У 401 юноши и 408 девушек в возрасте от 15 до 17 лет, обучающихся в колледжах г. Челябинска, проведены измерения роста и массы тела, а также опрос о питании по утвержденной методике, об уровне физической активности подростков, об уровне образования и ИМТ родителей, семейного дохода [22]. Определена распространенность дефицита массы тела, избыточной массы тела и ожирения по референсным кривым для детей и подростков 5–19 лет, предложенным ВОЗ [23]. Критерии Стьюдента, Манна – Уитни и Краскела – Уоллиса использовались для определения статистической значимости различий средних и медиан. Для оценки связи ИМТ подростков с показателями пищевой ценности рациона, уровнем физической активности и ИМТ родителей были построены модели множественной линейной регрессии с применением метода скользящего среднего.

Статистическую обработку данных проводили с использованием пакета прикладных статистических программ Statistica 10 (TIBCO Software Inc., Palo Alto, CA, USA). Статистическая значимость различий оценивалась на уровне значимости $\alpha=0,05$.

3. Результаты и обсуждение

Описательная статистика

Массив МБД настоящего исследования содержал переменные разного типа.

Количественные непрерывные переменные представлены такими показателями, как Возраст, Рост, Вес, ИМТ, Окружность талии (ОТ), Толщина кожной складки, Индекс курения, Пищевая ценность рациона (жиры, белки, углеводы, витамины и минеральные вещества). Показатели пищевой ценности рациона распределены по логнормальному закону. Остальные перечисленные показатели имели нормальное распределение.

Другая группа переменных – это *категоризованные переменные*. К ним относились категория ИМТ (нормальная масса, дефицит и выраженный дефицит массы, избыток и выраженный избыток массы), характеристики Кратности питания, Образование родителей и Материальная обеспеченность семьи. С точки зрения биостатистики это качественные (ранговые) переменные.

Дихотомические переменные, принимающие только два значения 0 и 1 (да/нет), представлены в нашем исследовании показателями Пол и Курение.

В качестве описательной статистики специалисты-медики традиционно используют табличную форму представления данных. Между тем идеально подходящей формой для быстрого визуального анализа информации справедливо считается графическое представление [24]. Более того, графическое представление данных предпочтительно, поскольку позволяет визуально выделить закономерности или аномалии в численных данных, а также дает первичную оценку набора данных для возможности применения в дальнейшем более сложных инструментов анализа. Фактически это те же данные, представленные в более наглядной, интуитивно понятной форме.

На рис. 1 представлено распределение юношей и девушек по категориям ИМТ с учетом возраста [23].

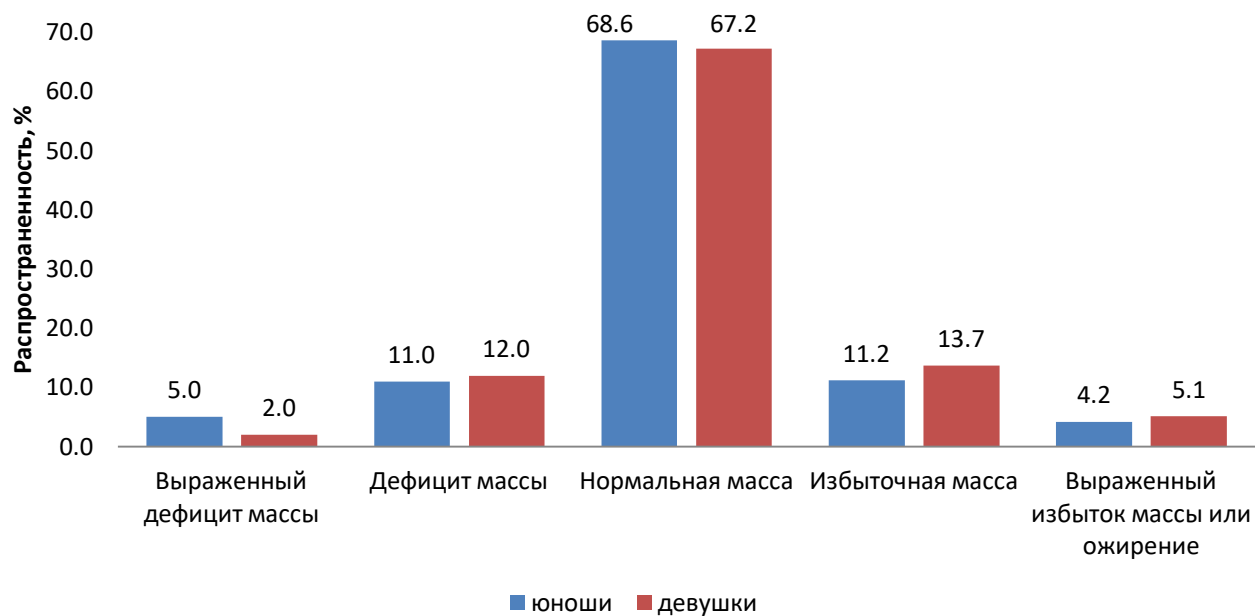


Рис. 1. Распределение юношей и девушек по категориям ИМТ

Около 67–68 % подростков имели нормальную массу тела, недостаточную или избыточную – около 11–13 %, и резко отличную от нормы ИМТ имеют около 5 % подростков. Распределение юношей и девушек по ИМТ с учетом возраста было близко к эталонной популяции ВОЗ [23]. Не было обнаружено статистически значимых различий между девушками и юношами в каждой из категорий ИМТ.

Визуализация связи ИМТ подростков с ИМТ родителей представлена на рис. 2.

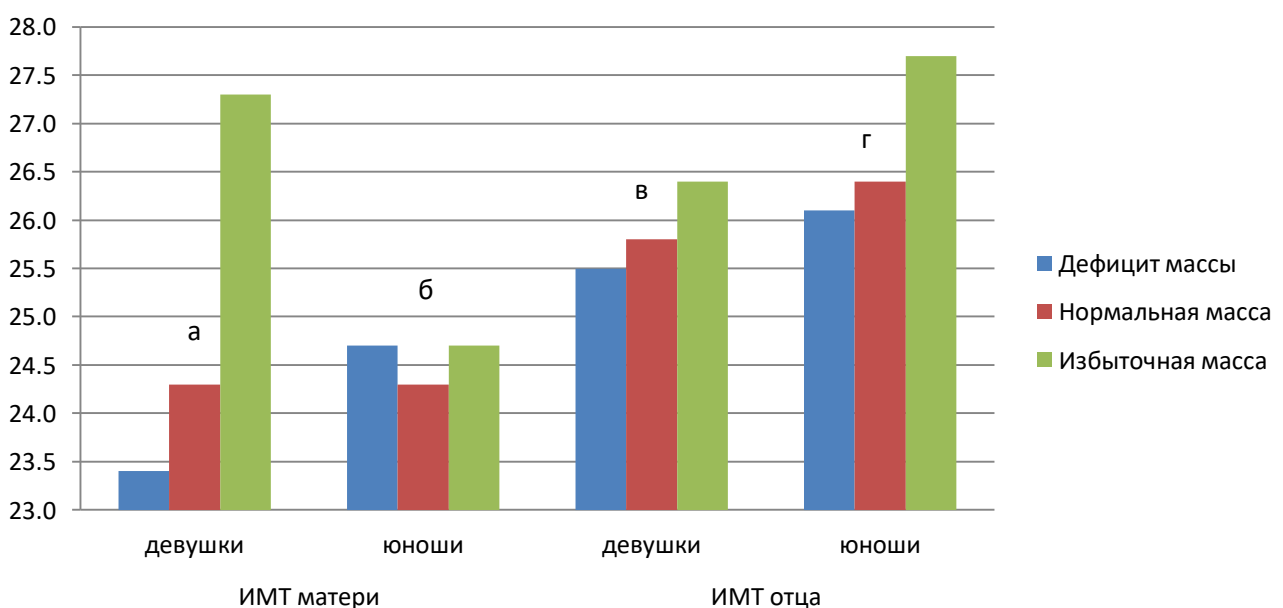


Рис. 2. Средние значения ИМТ родителей в группах подростков с дефицитом массы тела, нормальной массой тела и избыточной массой тела, кг/м²

В каждой из четырех групп (а – г) подростки разделены по трем категориям ИМТ (дефицит массы, нормальная и избыточная). Таким образом, левый столбец группы а показывает среднее значение ИМТ матерей (ИМТ_м), дочери которых имеют недостаточную массу тела (ИМТ_м = 23,4 кг/м²). Средний столбец группы а показывает среднее значение ИМТ матерей, у которых дочери имеют нормальную массу тела (ИМТ_м = 24,3 кг/м², т. е. несколько выше, чем в предыдущем случае, но различия не являются статистически значимыми (р=0,08)). Наконец, правый столбец показывает

среднее значение ИМТ_м, у которых дочери имеют избыточную массу тела. В этом столбце ИМТ_м = 27,3 кг/м², различия с группами слева являются статистически значимыми (p < 0,001). Таким образом, связь ИМТ девушек и ИМТ матерей существует. В то же время ИМТ юношей не имеет связи с ИМТ матери (группа б).

Группа в показывает средние значения ИМТ отцов для различных категорий ИМТ девушек. Статистически значимые различия выявлены для групп с дефицитом и избыточной массой тела (p=0,025). Группа г показывает средние значения ИМТ отцов для различных категорий юношей: у тучных юношей отцы имеют повышенный ИМТ. Статистически значимые различия выявлены в группах с дефицитом и избыточной массой тела (p=0,05).

Значимость различий оценивалась попарно, с использованием t-критерия Стьюдента, поскольку показатель ИМТ имеет нормальное распределение.

Таким образом, обнаружена статистически значимая связь ИМТ девушек и ИМТ матерей. Имеется связь ИМТ отцов с ИМТ подростков обоих полов, но только в группах с дефицитом и избыточной массой тела (p < 0,05). Наши результаты хорошо согласуются с выводами предыдущего исследования, где была обнаружена связь между ИМТ девочек и матерей и не обнаружена связь между ИМТ мальчиков и их матерей, ИМТ девочек или мальчиков и их отцов [25]. В другом исследовании, проанализировавшем данные порядка 100 000 детей и их родителей из шести стран, было обнаружено, что наследственная передача ИМТ наблюдалась во всех странах, независимо от их уровня экономического развития. В частности, оказалось, что у очень худых детей доля ИМТ родителей составляет около 20 % в целом, в то время как у детей, страдающих от ожирения, эта цифра достигает 60 % от ИМТ обоих родителей [26].

Рис. 3 демонстрирует отклонение от нормы в фактическом питании юношей и девушек.

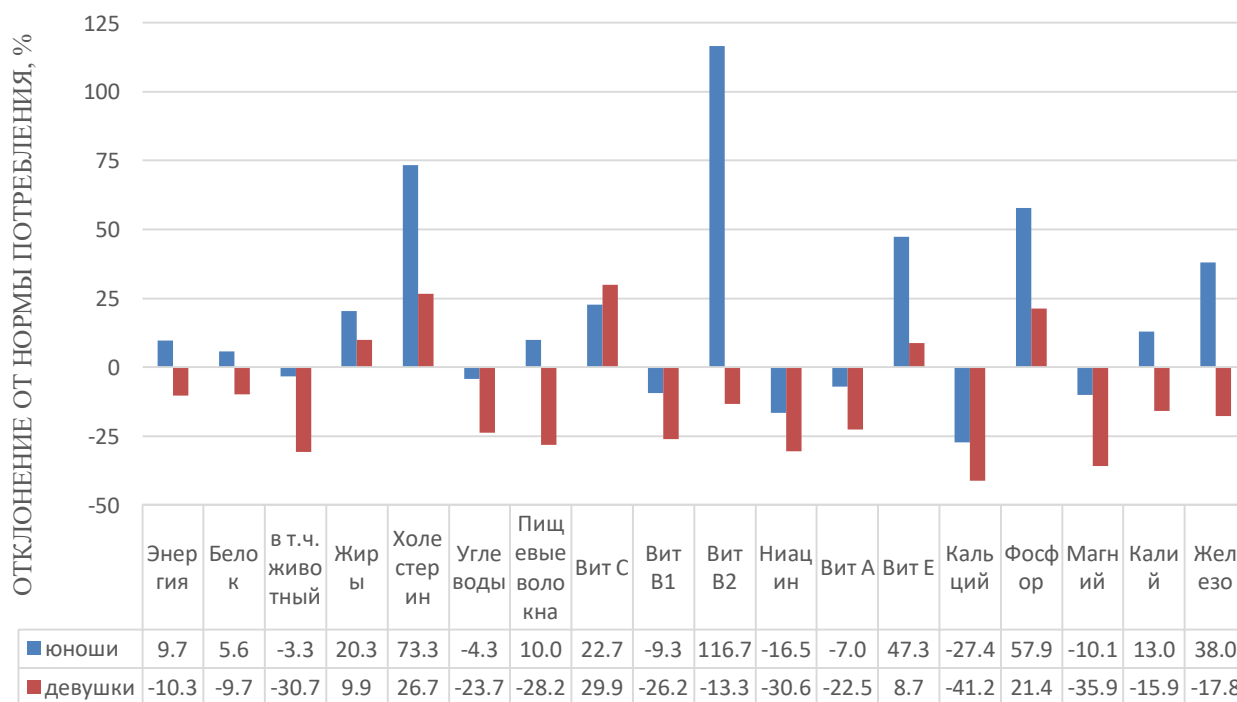


Рис. 3. Отклонение от нормы потребления пищевой ценности рациона, витаминов и минералов для юношей и девушек

Основное внимание следует обратить на показатели с наибольшим отклонением от нормы. За норму потребления на рис. 3 взято X=0. Наибольшие отклонения по избыточному потреблению (выше оси X) отмечены только для юношей по показателям: витамин В2, холестерин, фосфор, витамин Е, железо. Наибольшие

отклонения по недостаточному потреблению (ниже оси X) отмечены только для девушек по показателям: кальций, магний, белок животного происхождения, ниацин, пищевые волокна.

Выявлены значительные отличия в пищевой ценности рациона, потреблении витаминов и минералов между юношами и девушками. Причем как в случае разнонаправленных отклонений в потреблении (витамин B2, железо, пищевые волокна), так и однонаправленных (холестерин и витамин E).

Статистическое моделирование

При выборе метода построения модели, описывающей сопряженность ИМТ подростков с пищевыми и некоторыми непищевыми факторами, учитывался тип переменных и дизайн исследования (одномоментный срез). Предполагалось также, что каждое наблюдение является результатом независимых испытаний, образующих случайную выборку из генеральной совокупности.

В нашем случае зависимая переменная Y (ИМТ подростка) предположительно связана со многими независимыми переменными X_1, X_2, \dots, X_k (это пищевая ценность рациона, кратность питания в рабочие и выходные дни, коэффициент физической активности, ИМТ родителей, уровень образования родителей и т. д.). Y и X_k в нашем массиве данных являются непрерывными количественными переменными. К данным была применена процедура скользящего среднего для сглаживания случайных колебаний значений X и Y [27].

Все вышеизложенное определило выбор множественной линейной регрессии в качестве статистического метода для построения моделей.

Наиболее информативные предикторы были отобраны пошагово. Пошаговая процедура дает набор предикторов, которые обеспечивают приемлемое качество модели при минимальном количестве предикторов. Для учета взаимной коррелированности предикторов в модель были включены перекрестные члены.

Качество построенной модели характеризуется коэффициентом детерминации. Чем ближе R^2 к единице, тем меньше расчетные значения Y отличаются от экспериментальных.

Модели формирования ИМТ строились отдельно для юношей и девушек в силу различий в физиологии, пищевых паттернов, разного уровня физической активности и т. п.

Регрессионная модель взаимосвязи ИМТ юношей с показателями пищевой ценности, семейно-социальными и факторами генетической предрасположенности:

$$\text{ИМТ}_{\text{ю}} = -13,309 + [\text{Кол-во потребляемой пищи}] * 0,017 + [\text{Простые сахара}] * (-0,126) + [\text{ИМТ матери}] * 0,804 + [\text{Пищевые волокна}] * (-0,207) \quad (1)$$

$$R^2 = 0,86.$$

Регрессионная модель взаимосвязи ИМТ девушек с показателями пищевой ценности, семейно-социальными и генетически предрасположенными факторами:

$$\text{ИМТ}_{\text{д}} = 16,730 + [\text{ИМТ матери}] * 1,700 + [\text{Б:Ж}] * (-12,255) + [\text{Полисахариды}] * (-0,237) + [\text{Пищевые волокна} * \text{Полисахариды}] * 0,009 + [\text{Пищевые волокна}] * (-0,759) \quad (2)$$

$$R^2 = 0,91.$$

Коэффициенты регрессии перед всеми факторами, входящими в модель (1), и (2) являются статистически значимыми ($p < 0,01$).

Модель (2) имеет высокий коэффициент детерминации, $R^2 = 0,91$ (91 % дисперсии показателя ИМТ объясняется зависимостью от выбранных показателей, т. е. хорошо описывает фактические значения ИМТ).

На рис. 4 показана диаграмма рассеяния фактического/ожидаемого ИМТ девушек с учетом пищевой ценности и факторов семейной, социальной и генетической предрасположенности, которая является графическим представлением качества модели (2).

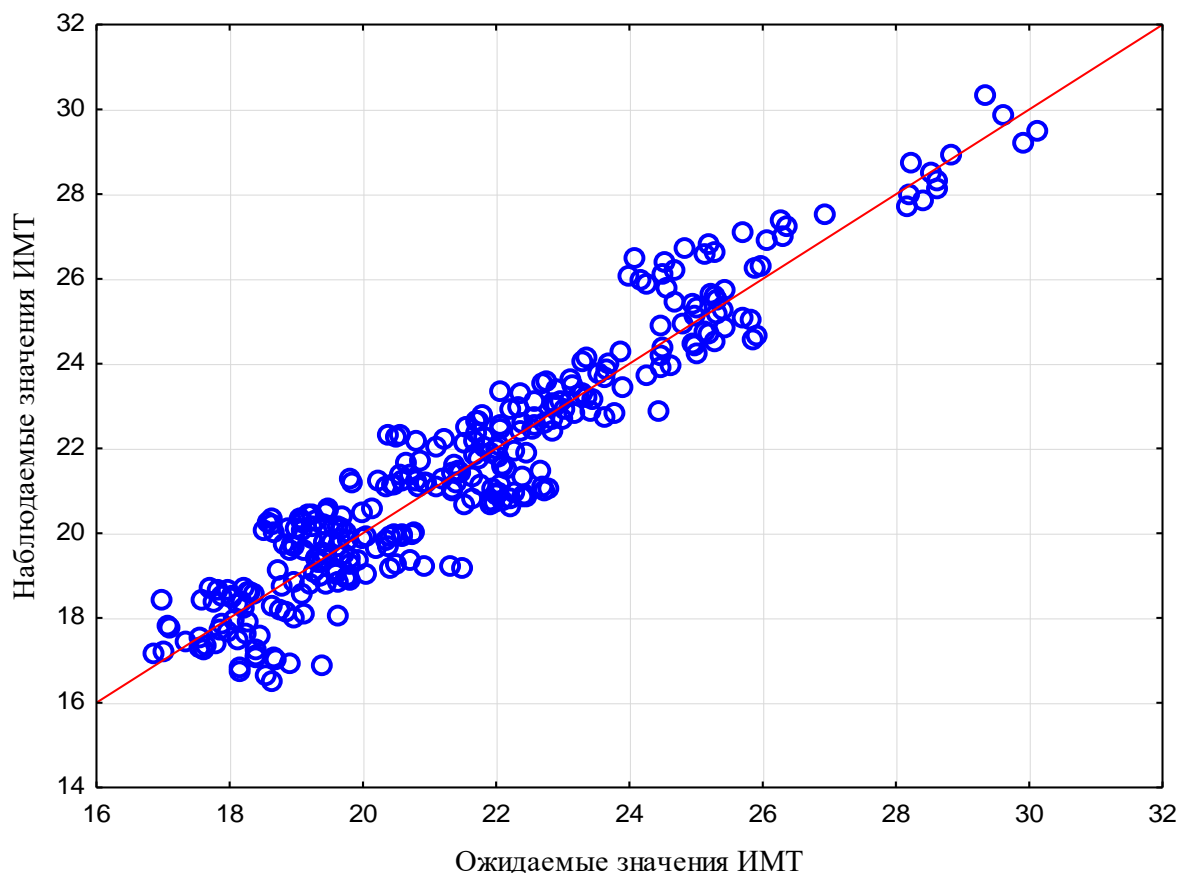


Рис. 4. Взаимосвязь наблюдаемых и рассчитанных по модели (2) значений ИМТ девушек

4. Заключение

Согласно принципу множественности моделей В. В. Налимова, «для объяснения и предсказания структуры и (или) поведения сложной системы возможно построение нескольких одинаково достоверных математических моделей» [28]. В настоящей работе с помощью комплекса статистических методов решена практическая задача эпидемиологии: построены модели, описывающие сопряженность ИМТ подростков с пищевыми и некоторыми непищевыми факторами; исследована распространенность дефицита массы тела, избыточной массы тела и ожирения среди подростков – учащихся колледжей. Результаты представлены в наглядной форме, доступной для специалистов – не математиков.

В предметном смысле результаты настоящего исследования позволяют сделать вывод о выявленных нарушениях режима и фактического питания подростков, что может негативно отразиться на их здоровье. Результаты исследования свидетельствуют о необходимости реализации профилактических мер по оптимизации питания в организованных коллективах и повышению уровня гигиенической грамотности обучающихся, их родителей, медицинских работников, педагогов и организаторов питания.

5. Список литературы

1. Огородникова, С. Ю. Методы визуализации данных в медико-биологических исследованиях / С. Ю. Огородникова, Е. Д. Константинова // Траектория исследований – человек, природа, технологии. – 2022. – № 3(3). – С. 4–18. – DOI 10.56564/27825264_2022_3_4. – EDN KBPYUR.
2. Гланц, С. Медико-биологическая статистика. – Москва : Практика. 1998. – 459 с. – ISBN 5-89816-009-4.
3. Shoukri , M. M. Statistical methods for health sciences / M. M. Shoukri, C. A. Pause. – 2nd Edition. – CRC Press, 1998. – 384 p. – ISBN 978-0849310959.
4. Красько, О. В. Статистический анализ данных в медицинских исследованиях – Минск : Международный государственный экологический университет им. А. Д. Сахарова, 2014. – 127 с. – ISBN 978-985-551-096-4.
5. Hosmer, D. Applied Logistic Regression / D. Hosmer, S. Lemeshow. – 2nd Edition. – New York : Wiley, 2000. – 397 p. – ISBN 0-471-35632-8.
6. Тьюки, Дж. Анализ результатов наблюдений. – Москва : Мир, 1981. – 695 с.
7. Лучинин, А. С. Прогностические модели в медицине / А. С. Лучинин // Клиническая онкогематология. Фундаментальные исследования и клиническая практика. – 2023. – Т. 16, № 1. – С. 27–36. – DOI 10.21320/2500-2139-2023-16-1-27-36. – EDN UZEQZU.
8. Кузнецова, А. В. Преодоление проблемы «черного ящика» при использовании методов машинного обучения в медицине / А. В. Кузнецова, О. В. Сенько, Ю. О. Кузнецова // Врач и информационные технологии. – 2018. – № S1. – С. 74–80. – EDN YMKAOL.
9. Приказ министерства труда и социального развития Российской Федерации № 768 от 28.12.2021.
10. Afifi, A. A. Statistical Analysis: A Computer Oriented Approach / A. A. Afifi, S. P. Azen. – New York : Academic Press, 1972. – 366 p. – ISBN: 9780120444502.
11. Afifi, A. A. Computer-aided multivariate analysis / A. A. Afifi, S. May, V. Clark. – 4th Edition. – Chapman&Hall/CRC, 2003. – 512 p. – ISBN 978-1584883081.
12. Mosteller, F. Data analysis and regression. A second course in statistics / F. Mosteller, J. W. Tukey. – Addison-Westly Publishing Company, 1978. – 317 p. – ISBN 978-0201048544.
13. Fletcher, R. H. Clinical epidemiology: The essentials / R. H. Fletcher, S. W. Fletcher. – 5th Edition. – Lippincott Williams & Wilkins, 2012. – 272 p. – ISBN 978-1451144475.
14. Вараксин, А. Н. Статистические модели регрессионного типа в экологии и медицине: монография / А. Н. Вараксин; Уральский гос. технический ун-т, Ин-т пром. экологии УрО РАН. – Екатеринбург : Гощицкий, 2006. – 255 с. – ISBN 5-98829-005-1. – EDN QLNFZT.
15. Распространенность избыточной массы тела и ожирения у детей / А. Н. Мартинчик, К. Э. Лайкам, Н. А. Козырева [и др.] // Вопросы питания. – 2022. – Т. 91, № 3(541). – С. 64–72. – DOI 10.33029/0042-8833-2022-91-3-64-72. – EDN VAQOJC.
16. Карлова, О. Б. Распространенность ожирения подростков в мире и Российской Федерации в 2012–2018 гг. / О. Б. Карлова, В. О. Щепин, А. А. Загоруйченко // Гигиена и санитария. – 2021. – Т. 100, № 4. – С. 365–372. – DOI 10.47470/0016-9900-2021-100-4-365-372. – EDN SRGLAX.
17. Vitamin D deficiency in childhood obesity is associated with high levels of circulating inflammatory mediators, and low insulin sensitivity / M. Reyman, A. A. Verrijn Stuart, M. van Summeren [et al.] // International Journal of Obesity. – 2014. – Vol. 38, No. 1. – P. 46–52. – DOI [10.1038/ijo.2013.75](https://doi.org/10.1038/ijo.2013.75).
18. Кинаш, М. И. Жирорастворимые витамины и иммунодефицитные состояния: механизмы влияния и возможности использования / М. И. Кинаш, О. Р. Боярчук // Вопросы питания. – 2020. – Т. 89, № 3. – С. 22–32. – DOI 10.24411/0042-8833-2020-10026. – EDN JZZZCV.
19. Пырьева, Е. А. Роль и место пищевых волокон в структуре питания населения / Е. А. Пырьева, А. И. Сафронова // Вопросы питания. – 2019. – Т. 88, № 6. – С. 5–11. – DOI 10.24411/0042-8833-2019-10059. – EDN TMNZUL.
20. Роль оптимального питания в профилактике сердечно-сосудистых заболеваний / Н. С. Сметнева, А. В. Погожева, Ю. Л. Васильев [и др.] // Вопросы питания. – 2020. – Т. 89, № 3. – С. 114–124. – DOI 10.24411/0042-8833-2020-10035. – EDN GCEZJJ.
21. Dietary Determinants of Metabolic Syndrome Parameters Differ by Gender in College Students / S. Saltzgeber, A. Nielson, H. Costello [et al.] // Nutrients. – 2019. – Vol. 11, No. 12. – DOI 10.3390/nu11122892.

22. Нутрициология и клиническая диетология: Национальное руководство / А. Л. Абалина, С. Е. Акользина, И. В. Аксенов [и др.]. – 2-е издание. – Москва : Общество с ограниченной ответственностью Издательская группа «ГЭОТАР-Медиа», 2021. – 1008 с. – (Национальные руководства). – ISBN 978-5-9704-6280-5. – DOI 10.33029/9704-6280-5-NKD-2021-1-1008. – EDN RLLIOP.
23. World Health Organization. *WHO child growth standards: training course on child growth assessment*, World Health Organization, 2021. Available online: <https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/bmi-for-age> (accessed on 8 August 2023).
24. Siegel, A. F. *Practical Business Statistics* / A. F. Siegel. – 7th Edition. – Academic Press, 2016. – 642 p. – ISBN 978-0128042502.
25. Parent-Offspring Associations in Body Composition: Findings From the Southampton Women's Survey Prospective Cohort Study / R. J. Moon, S. D'Angelo, C. R. Holroyd [et al.] // *The Journal of Clinical Endocrinology & Metabolism*. – 2023. – dgad128. – DOI [10.1210/clinem/dgad128](https://doi.org/10.1210/clinem/dgad128).
26. Dolton, P. The intergenerational transmission of body mass index across countries / P. Dolton, M. Xiao // *Economics & Human Biology*. – 2017. – Vol. 24. – P. 140–152. – DOI 10.1016/j.ehb.2016.11.005.
27. Применение методов скользящего среднего для построения регрессионных моделей в медико-экологических исследованиях / А. Н. Вараксин, Ю. В. Шалаумова, Т. А. Маслакова [и др.] // *Экологические системы и приборы*. – 2020. – № 6. – С. 12–21. – DOI 10.25791/esip.06.2020.1159. – EDN XTBFVAV.
28. *Налимов, В. В. Теория эксперимента*. – Москва : Наука, 1971. – 208 с.

Сведения об авторах:

Константинова Екатерина Даниловна, к. ф.-м. н., старший научный сотрудник, и. о. заведующего лабораторией математического моделирования в экологии и медицине Института промышленной экологии УрО РАН, г. Екатеринбург, Россия. Эл. почта: K.Konst@ecko.uran.ru

Маслакова Татьяна Анатольевна, к. ф.-м. н., научный сотрудник Института промышленной экологии УрО РАН, г. Екатеринбург, Россия.

Вараксин Анатолий Николаевич, д. ф.-м. н., профессор, главный научный сотрудник Института промышленной экологии УрО РАН, г. Екатеринбург, Россия.

Делец Светлана Семеновна, ассистент кафедры общей гигиены Южно-Уральского государственного медицинского университета Министерства здравоохранения Российской Федерации (ФГБОУ ВО ЮУГМУ Минздрава России), г. Челябинск, Россия.

APPROACHES TO HEALTH DATA STATISTICAL MODELING ON THE EXAMPLE OF ASSESSMENT OF THE ACTUAL NUTRITION OF ADOLESCENTS

E.D. Konstantinova¹, T.A. Maslakova¹, A.N. Varaksin¹, S.S. Delets²

¹ Institute of Industrial Ecology, Ekaterinburg, Russia

² Federal State Budgetary Educational Institution of Higher Education "South-Ural State Medical University" of the Ministry of Healthcare of the Russian Federation, Chelyabinsk, Russia

The issues of taking into account the peculiarities of working with biomedical data when solving epidemiological problems using statistical methods are discussed in this method paper. The importance of the correct choice of statistical methods for the analysis of MBD, as well as a visual graphical representation of experimental numerical results from the field of epidemiology, is emphasized. The MBD array was retrospectively analyzed, containing 809 cases and 185 variables of various types. As a method of analysis, multiple linear regression was chosen using a moving average approach. The practical problem of constructing models that describe the association of adolescent BMI with nutritional and some non-nutritive factors has been solved. The prevalence of underweight, overweight, and obesity among adolescent college students has been studied. The results of this study may be required in the development of preventive measures to optimize nutrition in organized groups, and will also help to increase the level of hygiene literacy of the interested circle of people.

Key words: statistical modeling; biomedical data; multiple linear regression; actual nutrition of adolescents.

References

1. Ogorodnikova, S. Y. Data visualization methods in biomedical research / S. Y. Ogorodnikova, E. D. Konstantinova // Research trajectory – man, nature, technology. – 2022. – Vol. 3, No. 3. – P. 4–18. – DOI 10.56564/27825264_2022_3_4. – EDN KBPYR. [in Russian].
2. Glantz S. Medico-biological statistics. – Moscow : Practice. 1998. – 459 p. – ISBN 5-89816-009-4 [in Russian].
3. Shoukri, M. M. Statistical methods for health sciences / M. M. Shoukri, C. A. Pause. – 2nd Edition. – CRC Press, 1998. – 384 p. – ISBN 978-0849310959.
4. Krasko, O. V. Statistical analysis of data in medical research – Minsk : International state ecological university named after A.D. Sakharov, 2014. – 127 p. – ISBN 978-985-551-096-4 [in Russian].
5. Hosmer, D. Applied Logistic Regression / D. Hosmer, S. Lemeshow. – 2nd Edition. – New York : Wiley, 2000. – 397 p. – ISBN 0-471-35632-8.
6. Tukey, J. Analysis of observational results. – Moscow : Mir, 1981. – 695 p. [in Russian]
7. Luchinin, A. S. Prognostic models in medicine / A. S. Luchinin // Clinical oncohematology. Basic research and clinical practice. – 2023. – Vol. 16, No. 1. – P. 27–36. – DOI 10.21320/2500-2139-2023-16-1-27-36. – EDN UZEQZU [in Russian].
8. Kuznetsova, A. V. Overcoming the "black box" problem when using machine learning methods in medicine / A. V. Kuznetsova, O. V. Senko, Yu. O. Kuznetsova // Doctor and information technology. – 2018. – No. S1. – P. 74–80. – EDN YMKAOL. [in Russian]
9. Order of the Ministry of Labor and Social Development of the Russian Federation No. 768 dated 12.28.2021 [in Russian].
10. Afifi A. A. Statistical Analysis: A Computer Oriented Approach / A. A. Afifi, S. P. Azen. – New York : Academic Press, 1972. – 366 p. – ISBN: 9780120444502.
11. Afifi, A. A. Computer-aided multivariate analysis / A. A. Afifi, S. May, V. Clark. – 4th Edition. – Chapman&Hall/CRC, 2003. – 512 p. – ISBN 978-1584883081.
12. Mosteller, F. Data analysis and regression. A second course in statistics / F. Mosteller, J. W. Tukey. – Addison-Westly Publishing Company, 1978. – 317 p. – ISBN 978-0201048544.

13. *Fletcher, R. H.* Clinical epidemiology: The essentials / R. H. Fletcher, S. W. Fletcher. – 5th Edition. – Lippincott Williams & Wilkins, 2012. – 272 p. – ISBN 978-1451144475.
14. *Varaksin, A. N.* Statistical models of regression type in ecology and medicine: monograph / A. N. Varaksin; Ural State Technical University, Institute of Industrial Ecology, Ural Branch of the Russian Academy of Sciences. – Ekaterinburg : Goshchitsky, 2006. – 255 p. – ISBN 5-98829-005-1. – EDN QLNFTZ [in Russian].
15. The prevalence of overweight and obesity in children / A. N. Martinchik, K. E. Laikam, N. A. Kozyreva [et al.] // *Problems of Nutrition*. – 2022. – Vol. 91, No. 3 (541). – P. 64–72. – DOI 10.33029/0042-8833-2022-91-3-64-72. – EDN VAQOJC [in Russian].
16. *Karpova, O. B.* The prevalence of adolescent obesity in the world and the Russian Federation in 2012-2018 / O. B. Karpova, V. O. Shchepin, A. A. Zagoruychenko // *Hygiene and Sanitation*. – 2021. – Vol. 100. – No. 4. – P. 365–372. – DOI 10.47470/0016-9900-2021-100-4-365-372. – EDN SRGLAX [in Russian].
17. Vitamin D deficiency in childhood obesity is associated with high levels of circulating inflammatory mediators, and low insulin sensitivity / M. Reyman, A. A. Verrijn Stuart, M. van Summeren [et al.] // *International Journal of Obesity*. – 2014. – Vol. 38, No. 1. – P. 46–52. – DOI [10.1038/ijo.2013.75](https://doi.org/10.1038/ijo.2013.75).
18. *Kinash, M. I.* Fat-soluble vitamins and immunodeficiency states: mechanisms of influence and possibilities of use / M. I. Kinash, O. R. Boyarchuk // *Problems of Nutrition*. – 2020. – Vol. 89, No. 3. – P. 22–32. – DOI 10.24411/0042-8833-2020-10026. – EDN JZZZCV [in Russian].
19. *Pyr'eva, E. A.* The role and place of dietary fiber in the structure of nutrition of the population / E. A. Pyr'eva, A. I. Safronova // *Problems of Nutrition*. – 2019. – Vol. 88, No. 6. – P. 5–11. – DOI 10.24411/0042-8833-2019-10059. – EDN TMNZUL, [in Russian].
20. The role of optimal nutrition in the prevention of cardiovascular diseases / N. S. Smetneva, A. V. Pogozeva, Yu. L. Vasiliev [et al.] // *Problems of Nutrition*. – 2020. – Vol. 89, No. 3. – P. 114–124. – DOI 10.24411/0042-8833-2020-10035. – EDN GCEZJJ [in Russian].
21. Dietary Determinants of Metabolic Syndrome Parameters Differ by Gender in College Students / S. Saltzgeber, A. Nielson, H. Costello [et al.] // *Nutrients*. – 2019. – Vol. 11, No. 12. – DOI 10.3390/nu11122892.
22. *Nutrition and clinical nutrition: national guidelines* / A. L. Abalina, S. E. Akolzina, I. V. Aksenov [et al.]. – 2th Edition. – Moscow : Limited Liability Company Publishing Group "GEOTAR-Media", 2021. – 1008 p. – (National guidelines). – ISBN 978-5-9704-6280-5. – DOI 10.33029/9704-6280-5-NKD-2021-1-1008. – EDN RLLIOP [in Russian].
23. World Health Organization. *WHO child growth standards: training course on child growth assessment*; World Health Organization, 2021. Available online: <https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/bmi-for-age> (accessed on 8 August 2023).
24. *Siegel, A.F.* Practical Business Statistics / A. F. Siegel. – 7th Edition. – Academic Press, 2016. – 642 p. – ISBN 978-0128042502.
25. Parent-Offspring Associations in Body Composition: Findings From the Southampton Women's Survey Prospective Cohort Study / R. J. Moon, S. D'Angelo, C. R. Holroyd [et al.] // *The Journal of Clinical Endocrinology & Metabolism*. – 2023. – dgad128. – DOI [10.1210/clinem/dgad128](https://doi.org/10.1210/clinem/dgad128).
26. *Dolton, P.* The intergenerational transmission of body mass index across countries / P. Dolton, M. Xiao // *Economics & Human Biology*. – 2017. – Vol. 24. – P. 140–152. – DOI 10.1016/j.ehb.2016.11.005.
27. Application of moving average methods for constructing regression models in medical and environmental studies / A. N. Varaksin, Yu. V. Shalaumova, T. A. Maslakova [et al.] // *Ecological systems and devices*. – 2020. – No. 6. – P. 12–21. – DOI 10.25791/esip.06.2020.1159. – EDN XTBFV [in Russian].
28. *Nalimov, V.V.* Theory of experiment. – Moscow : Nauka, 1971. – 208 p. [in Russian].