

**Антомонов М.Ю., Волощук Е.В.**

## **КОНСТРУИРОВАНИЕ ИНТЕГРАЛЬНЫХ ПОКАЗАТЕЛЕЙ КОЛИЧЕСТВЕННЫХ ПРИЗНАКОВ С ПОМОЩЬЮ ОДНОМЕРНЫХ И МНОГОМЕРНЫХ МЕТОДОВ СТАТИСТИКИ**

Описана информационная технология обработки первичных данных для одномерного, двухмерного и многомерного методов расчета интегрального показателя, а также представлены результаты оценки валидности разработанной технологии.

### **Введение**

Использование математической статистики стало уже традиционным в большинстве современных исследований: с одной стороны, для точной обработки результатов исследования, с другой — для создания математических моделей, отображающих процесс функционирования реальных систем или дающих им оценку.

Одной из интересных задач математической статистики является задача организации интегральных характеристик, которые позволяют обобщить множество зарегистрированных показателей в один с помощью определенных математических процедур. Особенно полезны такие характеристики в оценке биологических процессов и систем, поскольку они имеют несколько уровней организации: молекулярный, клеточный, органный, организменный. Поскольку одной из характеристик оптимального функционирования как организма в целом, так и его систем является понятие «здоровье», интегральные характеристики могут быть инструментом сравнения, анализа, прогнозирования состояния здоровья и основой для принятия решений по его сохранению и укреплению.

**Актуальность** разработки методов построения интегральных показателей (ИП) обусловлена отсутствием общепринятого алгоритма, который был бы одинаково эффективным для расчета интегральных характеристик здоровья на разных уровнях организации (как на индивидуальном, так и на популяционном) на основе разных показателей: антропометрических, функционального состояния, адаптационных возможностей организма, рождаемости, заболеваемости, смертности, инвалидности и т.п. [1–8]. Существующие модели разработаны преимущественно для решения узкого круга задач, их тип зависит от цели исследования и вида обрабатываемых данных.

Одной из таких задач, к примеру, является измерение и обработка антропометрических показателей. Вопреки существующему мнению, антропометрические показатели не только отражают телосложение человека и некие "эстетические" нормы, но также являются важными характеристиками прогнозирования риска развития различных заболеваний. Для оценки и объединения антропометрических показателей используются их соотношения, так называемые индексы физического развития. Преимущество индексов в этом конкретном случае состоит в их уникальности, информативности, наглядности, а также простоте получения и

применения. В то же время объединение таких оценок на более высоком уровне и их дальнейшее сопоставление не всегда математически корректно из-за разнообразия первичных показателей.

Указанную проблему можно решить с помощью информационных технологий. При этом становится возможной разработка подходов к расчету унифицированных оценок здоровья, элементами которых могут быть данные, имеющие разную значимость, вариабельность и размерность [9].

**Методы и объем исследования.** Результаты любого исследования в значительной степени зависят от формирования перечня информативных исходных показателей. Решающую роль в этой задаче играют экспертные оценки значимости и информативности отдельных показателей с учетом возможности их измерения в определенных шкалах, диапазона вариации, трудоемкости сбора информации. Отбор показателей для расчета интегральной оценки проводился в соответствии со следующими требованиями: информативностью, полнотой описания, уникальностью, оцениваемостью, репрезентативностью, избыточностью.

Важное значение для отбора количественных показателей имеет шкала их измерений. Наиболее информативными являются показатели, которые измеряются в шкале отношений, менее информативные — в ранговой и наименее информативные — в номинальной.

При расчете интегральных характеристик учитывалась направленность влияния отобранных показателей на исследуемый объект — положительное или отрицательное.

Иногда при первичном статистическом анализе обнаруживается избыточность данных. Это может потребовать привлечения дополнительных ресурсов для анализа информации по многим показателям или ухудшить результаты статистической обработки, если такие данные окажутся зашумляющими. В таких случаях целесообразно снизить размерность перечня исходных показателей с одновременным сохранением основной исходной информации.

В представленной работе размерность данных была сокращена путем привлечения суждений экспертов (оценивались значимость всех показателей и исключались показатели, значимость которых ниже заданного уровня), а также с помощью предварительного корреляционного анализа (оставлены по одному показателю от каждой зависимой пары).

Необходимо отметить, что решение о выборе метода сокращения размерности всегда базируется на априорном знании об особенностях решаемой задачи и ожидаемого результата, а также наличии или отсутствии временных и расчетных ресурсов.

Поскольку интегральный показатель по своей сути является математической сверткой первичных показателей с несколькими уровнями иерархии, для такого рода объединения показатели должны быть сопоставимы между собой, иметь одинаковые единицы и границы измерения. Чтобы привести первичные показатели к желаемому виду, необходимо их подвергнуть дополнительной математической обработке.

Расчет весовых коэффициентов может производиться как расчетными методами, так и с помощью экспертного мнения [10–11].

Для проверки практической пользы и значимости разработанной методики ее сравнивают с уже применяемой, валидность которой доказана. Для этого используют некий внешний критерий — показатели связи новой и старой методик.

Показатели валидности могут быть количественными и качественными. Для расчета количественного показателя (коэффициента валидности) сопоставляются результаты, полученные при применении разработанной методики с исходными данными или с данными родственных методик. В качестве количественных показателей могут быть использованы статистические критерии. Для определения степени валидности на практике наиболее часто применяются различные виды корреляционного анализа связи между результатами оцениваемой и эталонной методик. Оценки по критерию сравнения чаще бывают дихотомическими и ранговыми.

В зависимости от шкалы представленных показателей в сравниваемых рядах применяют те или иные методы корреляционного анализа и следующие показатели валидности:

- коэффициент ранговой корреляции Спирмена (ранги),
- коэффициент линейной корреляции Пирсона (отношения),
- коэффициент четырехклеточной сопряженности Пирсона (дихотомия),
- коэффициент точечно-бисериальной корреляции (сочетание дихотомической и интервальной шкал),
- коэффициент рангово-бисериальной корреляции (сочетание дихотомической и ранговой шкал).

При качественной оценке методики используют субъективное мнение исследователя или объективное оценивание специалистами (экспертами) в данной области, которое выражается в понятиях: «способна», «соответствует», «эффективна» или «не способна», «не соответствует», «не эффективна».

Предложенная технология была апробирована на результатах исследования уровня физического развития 1470 школьников (мальчиков и девочек) в возрасте от 6 до 17 лет, проведенного А.Г. Платоновой [12]. На основе антропометрических показателей — длина тела, масса тела, окружность грудной клетки, динамометрия правой и левой кисти, спирометрия — были рассчитаны индексы физического развития формулы (1)–(8) и интегральные показатели формулы (10)–(11).

Индекс Эрисмана:

$$\text{Инд}_\text{Э} = \frac{\text{ОГК} - \text{ДТ}}{2}, \quad (1)$$

где ОГК — окружность грудной клетки, ДТ — длина тела.

Индекс Бругша:

$$\text{Инд}_\text{Б} = \frac{\text{ОГК}}{\text{ДТ}} \times 100. \quad (2)$$

Индекс Вервека:

$$\text{Инд}_\text{В} = \frac{\text{ДТ}}{(2\text{МТ} + \text{ОГК})}, \quad (3)$$

где МТ — масса тела.

Индекс Кетле:

$$\text{Инд}_\text{К} = \frac{\text{МТ}}{0,0001\text{ДТ}^2}. \quad (4)$$

Индексы силы правой и левой кисти:

$$\text{Инд\_Спр} = \frac{D_{\text{пк}}}{MT}, \quad (5)$$

$$\text{Инд\_Слр} = \frac{D_{\text{лк}}}{MT}, \quad (6)$$

где  $D_{\text{пк}}$  — динамометрия правой кисти,  $D_{\text{лк}}$  — динамометрия левой кисти.

Жизненный индекс:

$$\text{Инд\_Ж} = \frac{\text{ЖЕЛ}}{MT}, \quad (7)$$

где  $S_{\text{пир}}$  — показатель спирометрии.

Индекс Пинье:

$$\text{Инд\_П} = DT - (MT + \text{ОГК}). \quad (8)$$

Наиболее важным понятием в биологии и медицине является понятие «норма» — оптимум функционирования и развития биологической системы, который встречается у большинства особей данного вида. Для биологических систем нормой очень часто служит среднее значение по какому-либо показателю, а слабое или сильное отклонение от среднего в ту или иную сторону является критерием определения патологии. Поэтому для избавления от размерности первичных показателей и установления для них единого диапазона измерения —  $[0;1]$  — была выбрана следующая формула:

$$d = \exp\left(-\frac{x_i - \bar{x}_i}{2\sigma^2}\right), \quad (9)$$

где  $\bar{x}$  — норма показателя (среднее значение),  $\sigma$  - стандартное отклонение по показателю.

Рассчитанные таким образом новые показатели называются нормированными эквивалентами и далее по тексту будут обозначаться буквой  $D$  (например,  $D_{\text{Э}}$  — нормированный индекс Эрисмана,  $D_{\text{МАС\_Т}}$  — нормированный показатель массы тела и т.д.)

В качестве показателя валидности методики были выбраны коэффициенты линейной корреляции Пирсона. Корреляции считались с помощью компьютерной программы Statistica v.8.0. Уровень значимости полученных коэффициентов корреляции принимался как достоверный при значениях  $p < 0,05$ .

**Содержание работы и обсуждение результатов.** Для получения интегрального показателя были применены одномерные, двухмерные и многомерные методы математической статистики.

В расчете ИП с помощью одномерных методов можно выделить основные этапы: формирование перечня информативных показателей, сокращение размерности данных (при необходимости), обезразмеривание показателей (приведение всех показателей к одной единице измерения), нормировка (установление жестких границ изменения значений показателей) и непосредственно расчет ИП.

Одномерные ИП целесообразно использовать в том случае, когда исходные показатели относительно независимы и имеют самостоятельную ценность.

В табл. 1 представлены результаты верификации одномерного способа расчета интегрального показателя. Были рассчитаны коэффициенты линейной корреляции между ИП, которые представляют собой среднее арифметическое и среднее геометрическое нормированных показателей физического развития (SR\_D\_A, SR\_D\_G) и индексами физического развития (жирным шрифтом обозначены достоверные коэффициенты корреляции). По результатам корреляционного анализа достоверная корреляция ИП наблюдается только с тремя индексами (ИНД\_В, ИНД\_К, ИНД\_П).

**Таблица 1**

*Коэффициенты корреляции интегральных показателей с индексами физического развития*

ИП	Стат. х-ки	Наименование индексов							
		Инд_Э	Инд_Б	Инд_В	Инд_К	Инд_Спр	Инд_Слв	Инд_Ж	Инд_П
SR_D_A	<i>r</i>	-0,15	-0,15	<b>0,21</b>	<b>-0,22</b>	0,03	-0,05	0,09	<b>0,20</b>
	<i>p</i>	<i>p</i> ≠,052	<i>p</i> ≠,056	<i>p</i> ≠,005	<i>p</i> ≠,003	<i>p</i> ≠,649	<i>p</i> ≠,544	<i>p</i> ≠,226	<i>p</i> ≠,008
SR_D_G	<i>r</i>	<b>-0,15</b>	<b>-0,15</b>	<b>0,21</b>	<b>-0,24</b>	0,05	-0,03	0,10	<b>0,21</b>
	<i>p</i>	<i>p</i> ≠,042	<i>p</i> ≠,044	<i>p</i> ≠,005	<i>p</i> ≠,002	<i>p</i> ≠,501	<i>p</i> ≠,654	<i>p</i> ≠,190	<i>p</i> ≠,005

В табл. 2 приведены коэффициенты корреляции между ИП (SR\_D\_A, SR\_D\_G) и нормированными эквивалентами индексов физического развития (D\_Э, D\_Б, D\_В и т.д.). Из таблицы видно, что увеличилось количество связей между нормированными индексами и ИП. Полученные результаты свидетельствуют о том, что нормировка повышает качество математической обработки данных.

**Таблица 2**

*Коэффициенты корреляции интегрального показателя с нормированными индексами физического развития*

ИП	Стат. х-ки	Наименование нормированных эквивалентов							
		D_Э	D_Б	D_В	D_К	D_Спр	D_Слв	D_Ж	D_П
SR_D_A	<i>r</i>	<b>0,232</b>	<b>0,239</b>	<b>0,341</b>	<b>0,129</b>	<b>0,268</b>	<b>0,385</b>	<b>0,162</b>	<b>0,198</b>
	<i>p</i>	0,002	0,002	< 0,001	0,09	< 0,001	< 0,001	0,033	0,009
SR_D_G	<i>r</i>	<b>0,227</b>	<b>0,233</b>	<b>0,316</b>	<b>0,136</b>	<b>0,292</b>	<b>0,425</b>	<b>0,162</b>	<b>0,218</b>
	<i>p</i>	0,003	0,002	< 0,001	0,074	< 0,001	< 0,001	0,033	0,004

В двумерном случае на первом этапе для определенных пар показателей в качестве формулы преобразования были выбраны отношения значений показателей с учетом априорной логики конструирования индексов физического развития. Например, если индекс Эрисмана рассчитывается с помощью показателей ОГК и ДТ (формула (1), то его аналогом будет отношение ОГК к ДТ. В результате такого преобразования появились новые характеристики, для которых были рассчитаны нормированные эквиваленты по формуле (9).

Двухмерные ИП можно использовать в случае, когда важны не столько сами исходные показатели, сколько их соотношения. Например, для определения уровня физического развития детей важны соотношения между антропометрическими показателями, т.е. гармоничность развития.

В табл. 3 представлены результаты верификации двухмерного способа расчета интегрального показателя. Показатели связи рассчитывались между нормированными аналогами индексов (AD\_Э, AD\_Б и т.д.) и нормированными индексами физического развития (D\_Э, D\_Б и т.д.).

Полученные результаты отображают довольно тесную связь между исследуемыми показателями (коэффициенты корреляции, размещенные по диагонали).

**Таблица 3**

*Коэффициенты корреляции аналогов индексов с нормированными индексами*

Нормир. индексы	Стат. х-ки	Наименования аналогов индексов							
		AD_Э	AD_Б	AD_В	AD_К	AD_Спр	AD_Слв	AD_Ж	AD_П
D_Э	<i>r</i>	<b>0,997</b>	<b>0,213</b>	0,016	-0,119	0,007	0,127	-0,086	0,016
	<i>p</i>	< 0,001	0,005	0,831	0,119	0,93	0,095	0,261	0,831
D_Б	<i>r</i>	1	<b>0,206</b>	0,03	<b>-0,13</b>	0,002	0,116	-0,1	0,03
	<i>p</i>	---	0,006	0,691	0,088	0,976	0,128	0,192	0,691
D_В	<i>r</i>	<b>0,314</b>	-0,072	<b>0,805</b>	<b>0,679</b>	-0,055	0,03	-0,005	<b>0,805</b>
	<i>p</i>	< 0,001	0,346	< 0,001	< 0,001	0,469	0,693	0,953	< 0,001
D_К	<i>r</i>	0,021	0,017	<b>0,546</b>	<b>0,821</b>	-0,053	-0,015	<b>0,153</b>	<b>0,546</b>
	<i>p</i>	0,788	0,823	< 0,001	< 0,001	0,488	0,847	0,044	< 0,001
D_Спр	<i>r</i>	0,002	-0,006	-0,103	-0,093	1	<b>0,527</b>	-0,026	-0,103
	<i>p</i>	0,976	0,933	0,177	0,222	---	< 0,001	0,737	0,177
D_Слв	<i>r</i>	0,116	0,048	-0,041	-0,052	<b>0,527</b>	1	<b>0,132</b>	-0,041
	<i>p</i>	0,128	0,535	0,59	0,498	< 0,001	---	0,084	0,59
D_Ж	<i>r</i>	-0,1	-0,058	0,012	0,122	-0,026	<b>0,132</b>	1	0,012
	<i>p</i>	0,192	0,449	0,875	0,111	0,737	0,084	---	0,875
D_П	<i>r</i>	<b>0,689</b>	<b>0,311</b>	<b>0,217</b>	<b>0,184</b>	0,016	0,12	-0,08	<b>0,217</b>
	<i>p</i>	< 0,001	< 0,001	0,004	0,015	0,83	0,116	0,299	0,004

Многомерные ИП могут рассчитываться на основе неограниченного количества показателей и с помощью многих методов математической статистики. Для начального преобразования были использованы метрика Брея-Кертиса и метрика Эвклида.

Метрика Брея-Кертиса:

$$\rho_{В\_К} = \frac{\sum_{i=1}^m |x_i - \bar{x}_i|}{\sum_{i=1}^m x_i + \sum_{i=1}^m \bar{x}_i}, \quad (10)$$

где  $\bar{x}_i$  — среднее арифметическое  $i$ -го показателя,  $m$  — количество показателей,  $m = 6$ .

Метрика Эвклида:

$$\rho_E = \sqrt{\sum_{i=1}^m (x_i - \bar{x}_i)^2}, \quad (11)$$

где  $\bar{x}_i$  — среднее арифметическое  $i$ -го показателя.

Для оценки валидности многомерного метода был рассчитан коэффициент линейной корреляции Пирсона между ИП, полученными по метриках Брея-Кертиса и Эвклида (БР\_КЕР, ЭВКЛ) с нормированными эквивалентами антропометрических показателей и индексов (табл. 4, 5).

**Таблица 4**

*Коэффициенты корреляции метрик и нормированных антропометрических показателей*

Метрики	Стат. х-ки	Наименование нормированных показателей							
		D_ДЛ_Т	D_МАС_Т	D_ОГК	D_Д_ПРК	D_Д_ЛВК	D_СПП	SR_D_A	SR_D_G
БР_КЕР	$r$	<b>-0,366</b>	<b>-0,42</b>	<b>-0,449</b>	<b>-0,482</b>	<b>-0,58</b>	<b>-0,395</b>	<b>-0,815</b>	<b>-0,83</b>
	$p$	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
ЭВКЛ	$r$	<b>-0,393</b>	<b>-0,382</b>	<b>-0,373</b>	<b>-0,45</b>	<b>-0,594</b>	<b>-0,399</b>	<b>-0,787</b>	<b>-0,845</b>
	$p$	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

**Таблица 5**

*Коэффициенты корреляции метрик и нормированных индексов*

Метрики	Стат. х-ки	Наименование нормированных индексов							
		D_Э	D_Б	D_В	D_К	D_СПП	D_СЛВ	D_Ж	D_П
БР_КЕР	$r$	<b>-0,258</b>	<b>-0,259</b>	<b>-0,33</b>	<b>-0,197</b>	<b>-0,163</b>	<b>-0,275</b>	-0,115	<b>-0,261</b>
	$p$	0,001	0,001	< 0,001	0,009	0,032	< 0,001	0,132	0,001
ЭВКЛ	$r$	<b>-0,198</b>	<b>-0,199</b>	<b>-0,273</b>	<b>-0,177</b>	<b>-0,187</b>	<b>-0,351</b>	<b>-0,152</b>	<b>-0,229</b>
	$p$	0,009	0,009	< 0,001	0,02	0,014	< 0,001	0,045	0,001

Отрицательный знак коэффициентов корреляции обусловлен способом нормировки, при котором большие значения показателей характеризуют лучшее состояние исследуемого объекта, в то время как метрики рассчитывают расстояние значений показателей от нормы, т.е. когда меньшие значения характеризуют лучшее состояние объекта. Из таблиц видно, что метрики с высокой достоверностью коррелируют с нормированными эквивалентами антропометрических показателей и индексов ( $p < 0,001$ ).

**Выводы.** На примере расчета ИП с помощью одномерного метода математической статистики было показано, что проведение нормировки значений первичных показателей может служить инструментом повышения их информативности и улучшения результата всей математической обработки данных.

Двумерный вариант построения интегрального показателя оптимально подходит для конструирования соотношений показателей, о чем

свидетельствует сильная корреляция двухмерного ИП с соответствующими нормированными эквивалентами индексов физического развития.

Многомерные ИП с высокой достоверностью коррелируют с нормированными эквивалентами и первичных показателей и индексов.

Таким образом, полученные результаты свидетельствуют о возможности расчета информативных интегральных показателей с помощью одно-, двух- и многомерных методов математической статистики.

1. *Гундаров И.А.* Оценка здоровья и влияющих на него факторов // Индивидуальная профилактика хронических неинфекционных заболеваний (пособие для врачей). — М.: ВНИИФК, 2001. — С. 5–20.
2. *Дергачева Л.Н.* Гигиенические аспекты формирования здоровья населения // Труды ИМКВЛ. — Владивосток: АГМА, 2003. — С. 38–45.
3. *Кику П.Ф., Веремчук Л.В.* Комплексная гигиеническая оценка среды обитания и здоровья населения Приморского края // Гигиена и санитария. — 2002. — №3. — С. 16–21.
4. *Сидоров А.А., Силич М.П.* Комплексная оценка демографического развития муниципального образования // Проблемы управления. — 2008. — № 1. — С. 29–35.
5. *Грановский Э.И., Снытин И.А.* Интегральные показатели загрязнения окружающей среды для оценки ее качества и состояния здоровья населения. Информационный листок — Алматы: КазГосИНТИ, 2009. — № 210. — 94 с.
6. *Белогокров В.П., Лозанский В.Р., Песина С.А.* Применение обобщенных показателей для оценки уровня загрязненных водных объектов // Комплексные оценки качества поверхностных вод. — Ленинград: Гидрометеиздат, 1984. — 34 с.
7. Гигиенические основы формирования перечней показателей для оценки и контроля безопасности питьевой воды / Г.Н. Красовский, Ю.А. Рахманин, Н.А. Егорова, А.Г. Малышева и др. // Гигиена и санитария. — 2010. — № 4. — С. 8–12.
8. *Щербо А.П.* Окружающая среда и здоровье: подходы к оценке риска. — СПб.: СПбМАПО, 2002. — 376 с.
9. *Антомонов М.Ю.* Математическая обработка и анализ медико-биологических данных. — Киев: НИИ ПВМ ВС Украины, 2006. — 558 с.
10. *Бешилев С.Д., Гурвич Ф.Г.* Математико-статистические методы экспертных оценок. — 2-е изд., перераб. и доп. — М.: Статистика, 1980. — 43 с.
11. *Голуб Дж., Ван Лоун Ч.* Матричные вычисления. — М.: Мир, 1999. — 156 с.
12. *Кучма В.Р., Платонова А.Г., Скоблина Н.А.* Физическое развитие московских и киевских школьников // Гигиена и санитария. — 2011. — № 1. — С. 75–78.

ГУ “Институт гигиены и медицинской  
Экологии имени А.Н. Марзеева  
НАМН Украины”, Киев

Получено 21.05.2012