32. Keul, J. Adaptation to training and Performance in Elite Athletes / J. Keul et al. // Reserch Quarterly for Exersice and Sport. – 1996. – V. 67. – Suppl. № 3. – P. 29–36.

33. Faude. O. Lactate Threshold Concepts: How Valid are They? / O. Faude. W. Kindermann. T. Meyer // Sports Med. – 2009. – V. 39. – № 6. – P. 469–490.

34. Gladden L. B. Current Trends in Lactate Metabolism: Introduction. / L. B. Gladden // Medicine & Science in Sports & Exercise. – 2008. – V. 40. – № 3. – P. 475–476.

35. Noakes T. D. From catastrophe to complexity: a novel model of integrative central neural regulation of effort and fatigue during exercise in humans: summary and conclusions / T. D. Noakes. A. St. C. Gibson. E.V. Lambert // Br. J. Sports. Med. – 2005. – V. 39. – P. 120–124.

36. Borg, G. A. Psychophysical bases of perceived exertion / G. A. Borg // Medicine and Science in Sports Exercise. – 1982. – V. 14. – № 5. – P. 377–381.

Поступила 14.03.2012

ПРОБЛЕМА АППРОКСИМАЦИИ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

В.Е. Ягур, канд. мед. наук, доцент,
Белорусский государственный медицинский университет.
Ю.М. Досин, д-р мед. наук, профессор,
М.В. Пуренок, канд. мед. наук,
Белорусский государственный университет физической культуры

В статье проведена аппроксимация эмпирических распределений трех антропометрических параметров (длина тела. масса тела. индекс Кетле) в соответствии с теорией обобщенных (универсальных) распределений. Показана возможность вычислять выравнивающее теоретическое распределение. определять ключевые параметры аппроксимирующего распределения (мода. точки перегиба) и визуализировать аппроксимирующее распределение.

In accordance with the theory of generalized (universal) distributions an approximation of empirical distributions of the three anthropometric parameters (body length, body mass, and Quetelet index) is carried out in the paper. A possibility to calculate an equalizing theoretical distribution, to determine the key parameters of the approximating distribution (mode. inflection points). and to visualize the approximating distribution is demonstrated.

При статистическом анализе биомедицинских данных важной является проблема нахождения подходящего закона распределения для описания группированного вариационного ряда. Выборочные характеристики – среднее арифметическое значение выборки и показатели ее варьирования (дисперсия. стандарт-

ное отклонение, коэффициент вариации) – не содержат полной информации о законе распределения генеральной совокупности.

Зачастую невозможно судить о законе распределения по эмпирической вариационной кривой, поскольку на ней разнообразие случайных причин. Между тем закон распределения биомедицинских признаков дает возможность избежать ошибок в оценке генеральных параметров на основании выборочных показателей и позволяет использовать адекватные методы их статистического анализа [2].

Чаще результаты измерений количественных характеристик биомедицинских объектов в шкале интервалов или отношений существенно отличаются от нормального (гауссового) распределения [8, 9, 11]. Лишь 10–20 % распределений количественных признаков, встречающихся в биомедицинских исследованиях, являются приближенно нормальными. Это означает, что применение параметрических методов статистики (критерий Стьюдента, дисперсионный, регрессионный, факторный анализы) в большинстве случаев не является обоснованным. Для указанных видов анализа параметрические методы могут применяться в том случае, когда все анализируемые одновременно признаки имеют нормальное распределение, здесь требуются непараметрические методы статистики или иные подходы [9].

Для надежного установления нормальности распределения требуется большое число наблюдений (порядка тысяч) или проверка статистических гипотез независимости и одинаковой распределенности с помощью соответствующих непараметрических критериев согласия – Смирнова, Колмогорова, Лемана-Розеблатта (омега-квадрат), критерий согласия Пирсона χ^2 и т. д. Критерий Стьюдента может использоваться для проверки нулевой гипотезы (H_0) об однородности математических ожиданий при условии: 1) результаты наблюдений имеют нормальные распределения; 2) дисперсии в двух выборках совпадают [10].

В прикладной статистике имеется два подхода к обработке исходных данных – детерминированный и модельно-вероятностный. При детерминированном подходе данные анализируются сами по себе без нужды в знании характера распределения данных, но тогда невозможно оценить погрешность рассчитанных показателей.

Цель научного исследования заключается в выявлении закономерностей на выборке и дальнейшей экстраполяции полученных результатов на всю генеральную совокупность (популяцию). Для этого и используется модельно-вероятностный подход, в основе алгоритма расчета которого лежит вероятностная модель порождения данных (модель случайной выборки).

Существуют различные параметрические семейства распределений числовых случайных величин (семейства нормальных распределений, логарифмически нормальных, экспоненциальных, гамма-распределений), зависящих от нескольких параметров, достаточно полно описывающих распределение.

Другой подход заключается в применении методов преобразования, трансформации данных имеющегося эмпирического распределения в нормальное с

помощью логарифмирования (без сдвига и со сдвигом), извлечения квадратного корня, обратного преобразования и т. д. Такая трансформация может оказаться удачной в плане приближения эмпирических данных к нормальному распределению, но следует помнить, что преобразование данных ведет к изменению единиц их измерения и потере ими своего физического смысла. поэтому результаты анализа преобразованных данных трудно оценивать [1].

Третий подход определения характера эмпирического распределения предложен доктором технических наук В.В. Нешитым [3–7]. Им разработана теория обобщенных (универсальных) распределений, которая включает три системы непрерывных распределений, систему дискретных распределений, взаимосвязанную с системой кривых роста новых событий, методы установления типа выравнивающей кривой и нахождения оценок параметров по номограммам. Разработана серия компьютерных программ для работы с указанными системами.

Обычно в качестве основной вероятностной модели используется нормальный закон распределения, который не позволяет с достаточной точностью выровнять (аппроксимировать) эмпирические распределения, зависящие от случайных величин. Для выравнивания большого разнообразия статистических распределений В.В. Нешитым построены три системы непрерывных распределений, задающихся одной, двумя или тремя плотностями.

Первая система непрерывных распределений используется для выравнивания и прогнозирования статистических распределений таких случайных величин, последующие значения которых образуются из предыдущих путем их изменения (сдвига) на некоторую постоянную величину. При этом форма кривой распределения не изменяется. Не изменяются также центральные моменты, но изменяется среднее значение и коэффициент вариации. Средние значения таких случайных величин растут во времени по линейному закону, но могут иметь и другой закон роста. Частным случаем первой системы непрерывных распределений является нормальный закон.

Вторая система непрерывных распределений используется для выравнивания и прогнозирования статистических распределений таких неотрицательных случайных величин, последующие значения которых образуются из предыдущих путем их умножения на некоторую постоянную величину. При этом форма кривой распределения меняется, но неизменными остаются коэффициент вариации и параметры k, u. Средние значения логарифмов таких случайных величин изменяются во времени по линейному закону, а сами случайные величины – по показательному закону. Частным случаем второй системы непрерывных распределений является логарифмически нормальный закон.

Третья система непрерывных распределений используется для выравнивания и прогнозирования статистических распределений таких неотрицательных случайных величин, последующие значения которых образуются из предыдущих путем их возведения в некоторую степень. Средние значения двойных логарифмов таких величин изменяются во времени по линейному закону, а сами случайные величины – по двойному показательному закону. Частным случаем

третьей системы непрерывных распределений является двойное логарифмически нормальное распределение, т. е. случайная величина «lnlnY» распределена по нормальному закону.

Теория обобщенных распределений позволяет вычислять минимально необходимый объем выборки (n), при котором распределение среднего арифметического n независимых одинаково распределенных случайных величин можно считать нормальным. Кроме того, можно рассчитать доверительный интервал для среднего при заданной доверительной вероятности и объеме выборки при любом законе распределения случайной величины, рассчитать коэффициенты асимметрии и эксцесса.

Материал и методы. Рассмотрим возможности оценивания параметров распределения биомедицинских данных на примере анализа антропометрических признаков: длины тела (ДТ), массы тела (МТ), индекса Кетле (ИК) у мужчин в возрасте от 18 до 36 лет (n=316). Данная выборка представляет собой фрагмент материалов по изучению индивидуальных и популяционных особенностей физического типа белорусов, собранных И.И. Саливоном во время комплексных экспедиций 1975–1985 гг., организованных отделом антропологии и экологии Института искусствоведения, этнографии и фольклора НАН Б в различных регионах Беларуси [12].

Избранные для анализа антропометрические признаки были аппроксимированы с помощью традиционного подхода – нормальным и логнормальным распределениями, а также с помощью второй системы непрерывных распределений по В.В.Нешитому, которая задается тремя обобщенными плотностями [3–6]. Вычисление аппроксимирующих распределений по указанным выше формулам проведено с помощью компьютерной программы SNR2V08A, разработанной В.В. Нешитым [7].

Подбор закона распределения для антропометрических признаков (длина тела, масса тела, индекс Кетле, далее – ДТ, МТ, ИК), а также построение гистограмм их эмпирических распределений и аппроксимирующих кривых в соответствии с нормальным и логнормальным распределениями, а также рядом других теоретических распределений, осуществлялся с помощью программы Statistica 6.0. Оценка статистической значимости различий между фактическими и ожидаемыми значениями изученных признаков проводилась с помощью критерия согласия Пирсона χ^2 [11].

Результаты и обсуждение. Традиционные расчеты средних значений (mean) и стандартных отклонений (SD) ДТ, МТ и ИК представлены в таблица 1. Кроме того, в этой же таблице приведены другие структурные средние: медиана (median), минимум (min), максимум (max), нижний (LQ_{25}) и верхний (UQ_{75}) квартили. Гистограммы эмпирических распределений признаков «длина тела», «масса тела», «индекс Кетле», а также их аппроксимации с помощью кривой нормального закона распределения представлены на рисунки 1–3, а на рисунке 4 – аппроксимация признака «индекс Кетле» с помощью кривой логнормального закона распределения.

Таблица 1 – Длина и масса тела, индекс Кетле мужчин от 18 до 36 лет (n=316)

Параметр	Возраст, лет	Длина, м	Масса тела, кг	ИК, кг/м²
MEAN	26,9	1,734	71,3	23,7
MEDIAN	27	1,733	69,9	23,2
SD	4,4	0,06	9,6	2,9
MIN	17,0	1,580	50	17,6
MAX	35,0	1,905	102	34,7
LQ_{25}	24,0	1,695	64,6	21,8
UQ_{75}	30,0	1,772	76,9	25,2

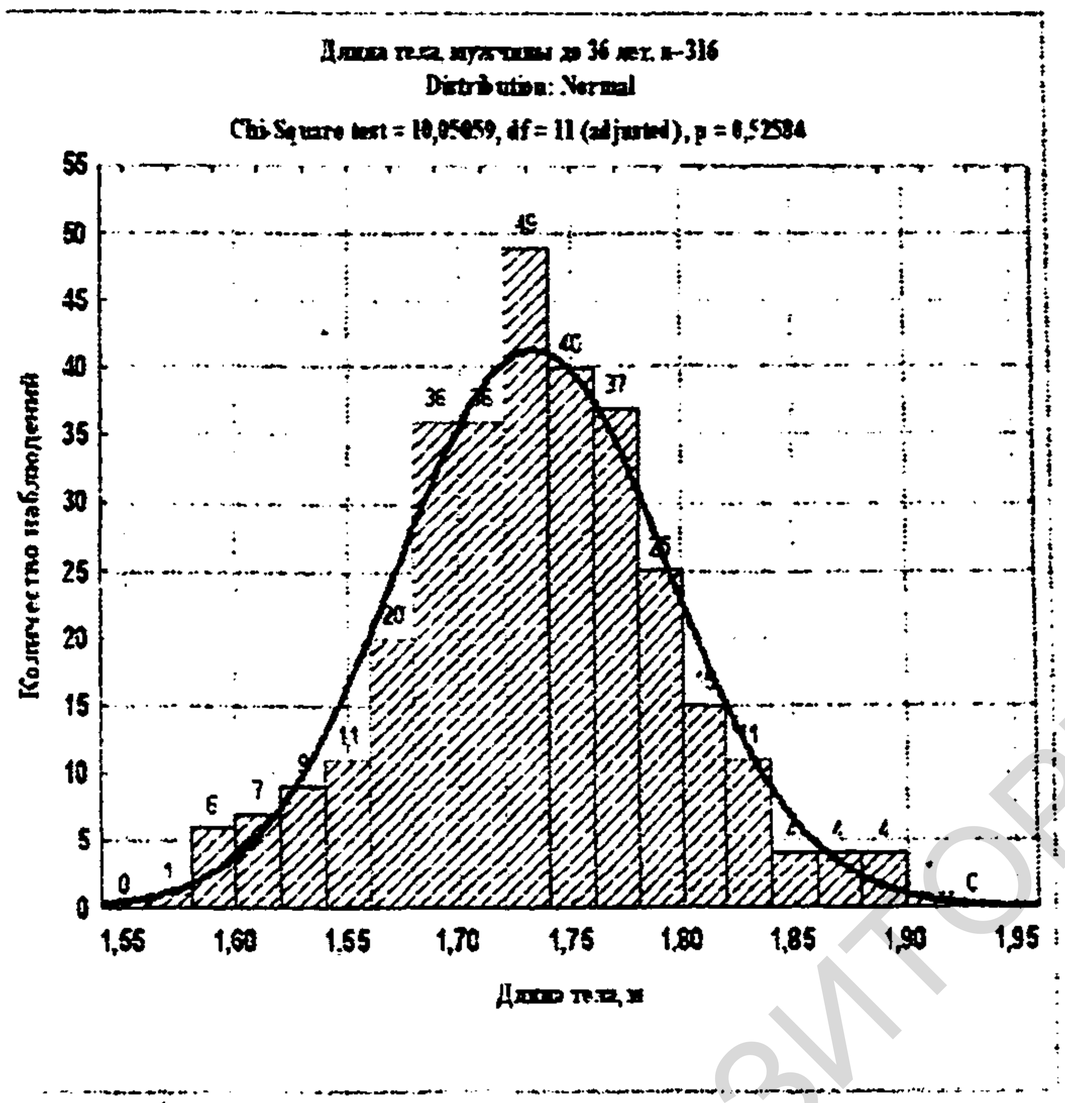

Рисунок 1 – Гистограмма эмпирического распределения признака «длина тела» и кривая ее аппроксимации нормальным распределением

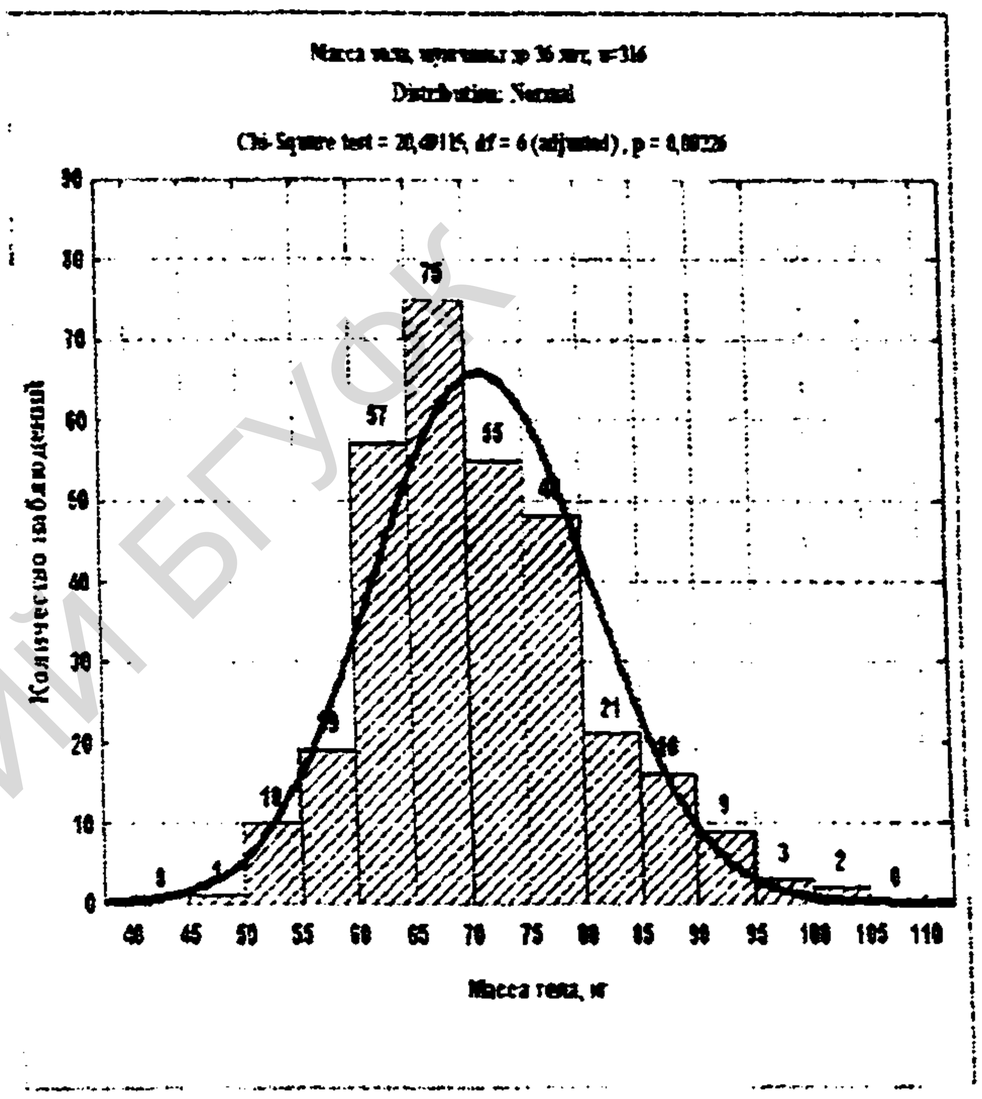

Рисунок 2 – Гистограмма эмпирического распределения признака «масса тела» и кривая ее аппроксимации нормальным распределением

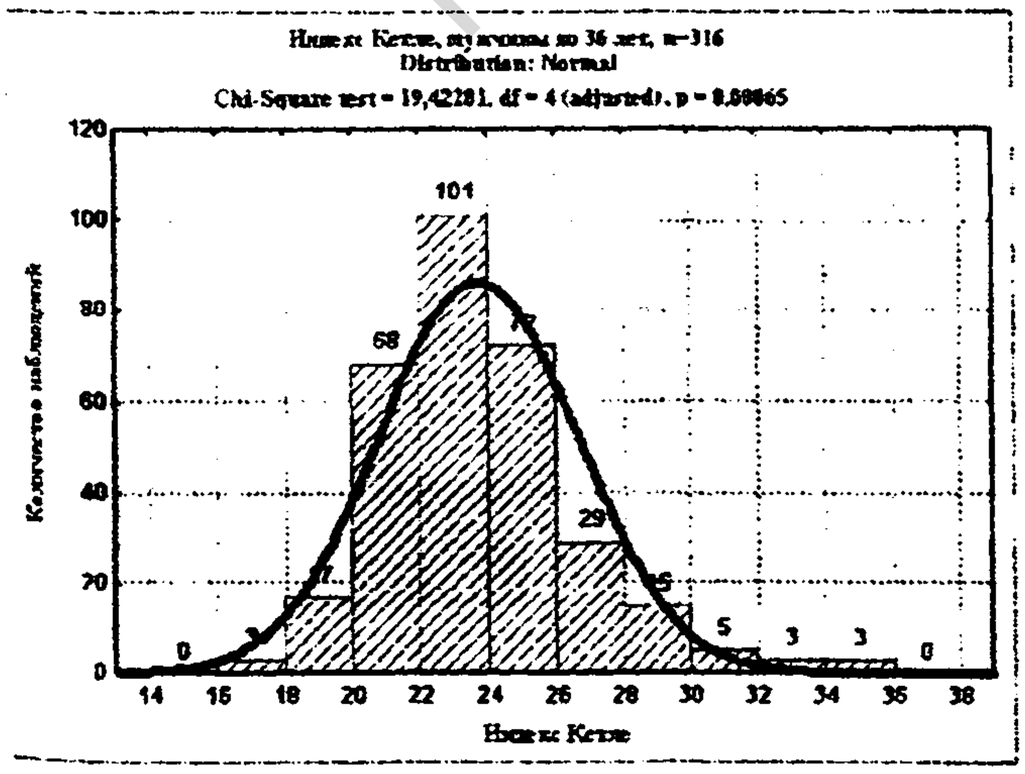

Рисунок 3 – Гистограмма эмпирического распределения признака «индекс Кетле» и кривая ее аппроксимации нормальным распределением

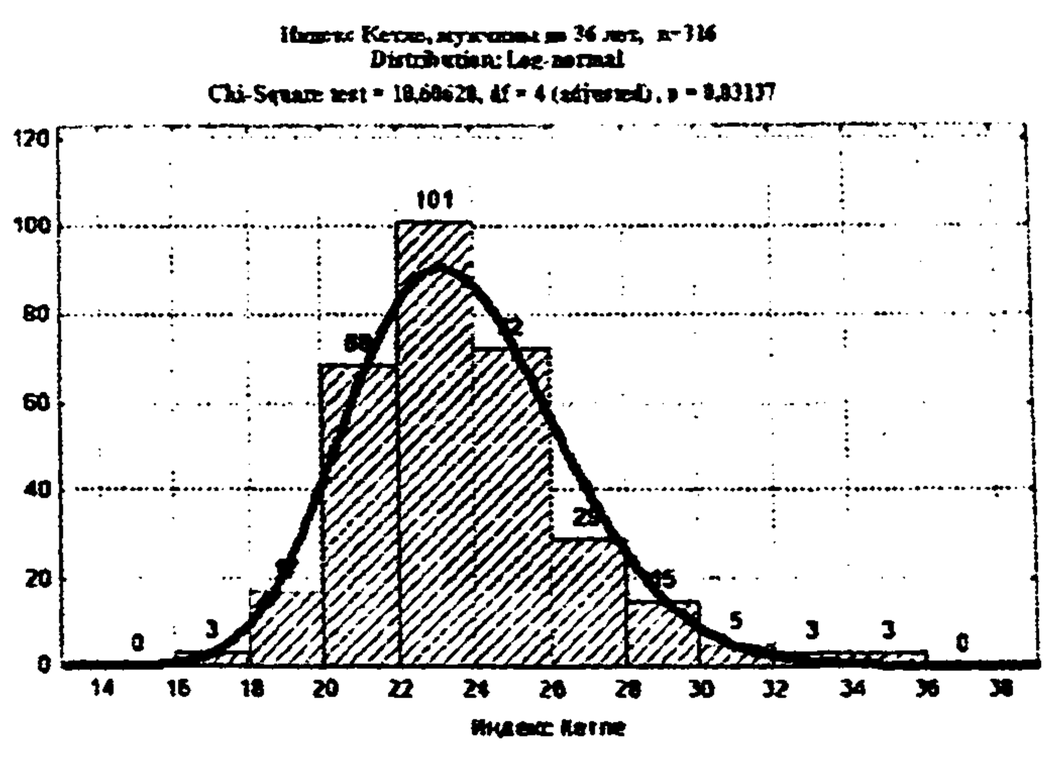

Рисунок 4 – Гистограмма эмпирического распределения признака «индекс Кетле» и кривая ее аппроксимации логнормальным распределением

Гистограмма эмпирического распределения признака «длина тела» соответствует теоретическому нормальному закону (χ^2=10,1; p=0,53).

Признак «масса тела» не аппроксимируется нормальным распределением: χ^2=20,5;,p=0,002). Это означает, что нулевая гипотеза (H_0) о соответствии эмпирического распределения признака «масса тела» нормальному закону должна быть отвергнута и принята альтернативная гипотеза (H_1).

Аппроксимация эмпирического распределения признака «масса тела» логнормальным распределением позволяет принять H_0 (χ^2=9,0; p=0,17), так как p>0,05.

Аппроксимация признака «индекс Кетле» нормальным и логнормальным распределениями показала, что этот признак не аппроксимируется ни нормальным (χ^2=19,4; p=0,0007), ни логнормальным распределениями (χ^2=10,6; p=0,03).

В соответствии с теорией обобщенных распределений эмпирическое распределение признака «длина тела» хорошо аппроксимируется теоретическим распределением 3-го типа (рисунок 5), параметры которого указаны в таблице 2.

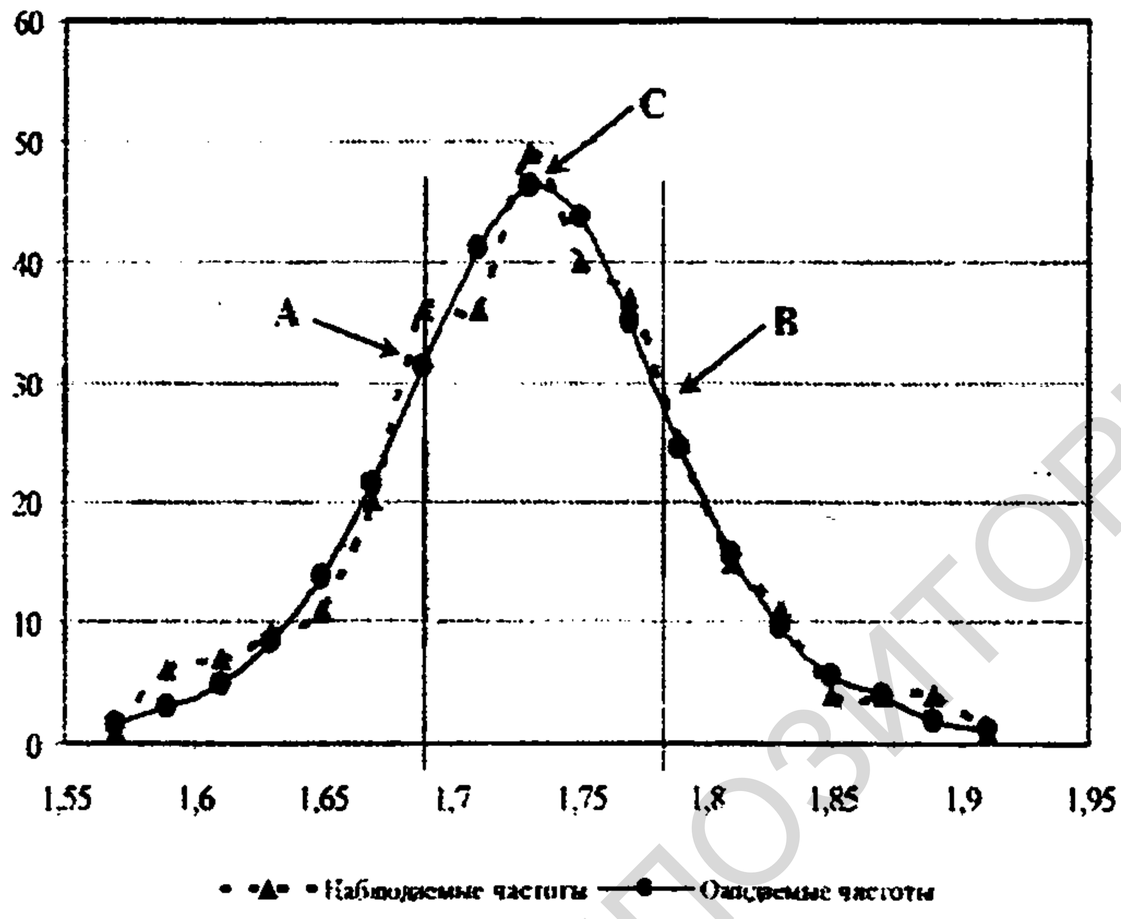

Рисунок 5 – Аппроксимация эмпирического распределения признака «длина тела» теоретическим распределением третьего типа (χ^2=4,365; p=0,9963)

Таблица 2 – Параметры третьего типа распределения признака «длина тела»

Параметр	Значение
αu	–1.260655754 E-14
β	57,85987
γ	44,44389
k = γ / β	0,76813
u	–1.568881
N	8,31809808 E-10
C (мода). м	1,734
A. м	1,690
B. м	1,777

Параметры C (мода), A и B (точки перегиба, т. е. точки, в которых вторая производная от плотности распределения равна 0) позволяют математически строго выделить центральный (от A до B) и периферические диапазоны (<A и >B) и присвоить им балловые значения 2 (от A до B), 1 (<A) и 3 (>B). Такая градация диапазонов длины тела позволяет использовать данный признак для выделения соматотипов мужчин на основе полуколичественной (балловой) оценки антропометрических данных. Судя по уровню χ^2, использование теории обобщенных распределений позволило более точно определить характер распределения признака «длина тела» у мужчин в возрасте от 18 до 36 лет.

Аналогичный подход был применен для аппроксимации эмпирического распределения признака «масса тела», который, как было показано, нельзя аппроксимировать с помощью нормального распределения, а также признака

«индекс Кетле», не поддающийся выравниванию не только нормальным, но и логнормальным распределением.

В рамках теории обобщенных распределений эмпирическое распределение признака «масса тела» хорошо аппроксимируется теоретическим распределением 3-го типа (рисунок 6), параметры которого указаны в таблице 3.

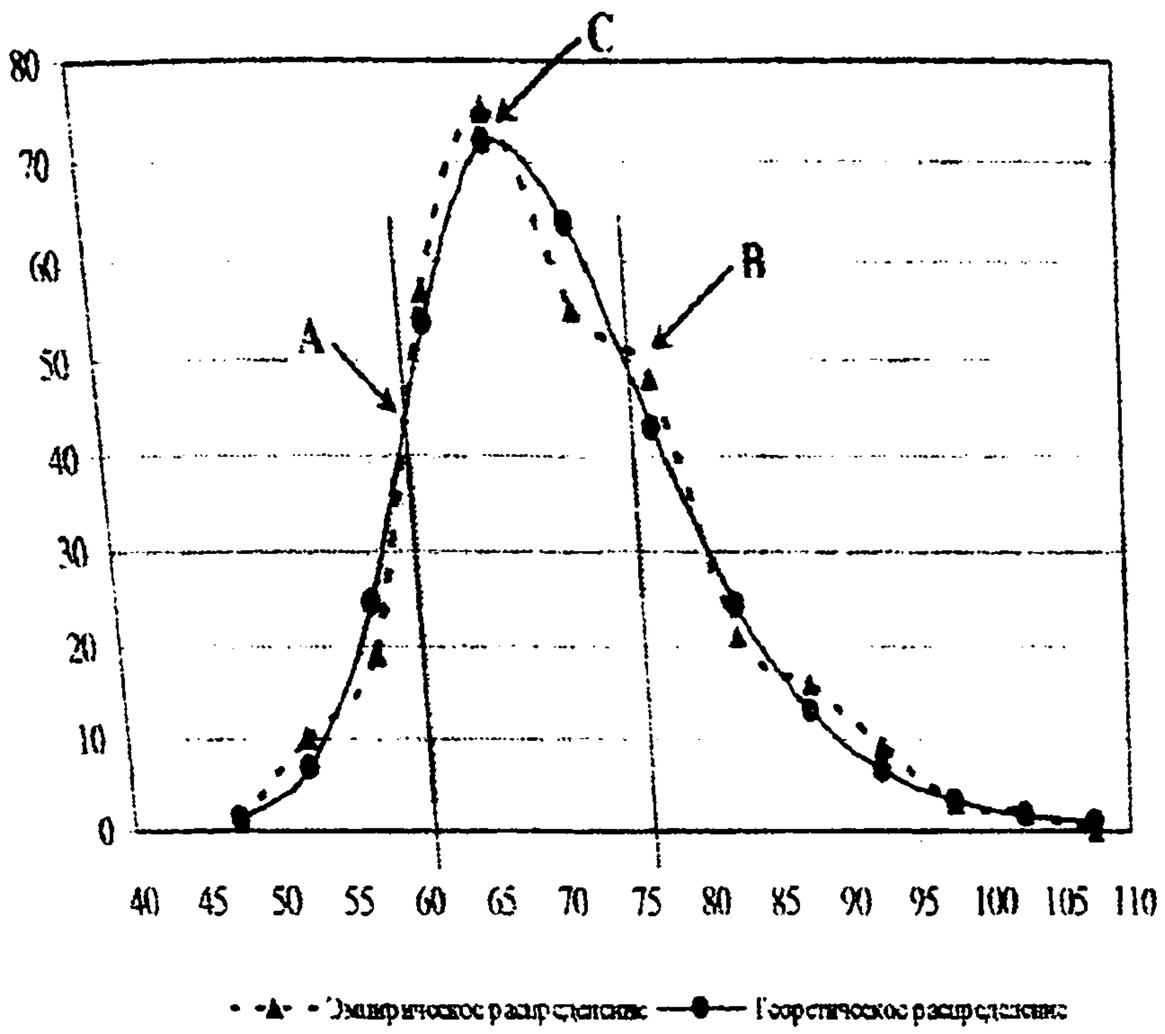

Рисунок 6 – Аппроксимация эмпирического распределения признака «масса тела» теоретическим распределением третьего типа (χ^2=6,856; p=0,9963)

Таблица 3 – Параметры третьего типа распределения признака «масса тела»

Параметр	Значение
αu	–1.40898714 E-16
β	8,723789
γ	21,22715
k = γ / β	2.4346
u	–0.3389107
N	1,17707527 E-37
C (мода). кг	68,3
A. кг	60.6
B. кг	75.9

Эмпирическое распределение признака «индекс Кетле» прекрасно аппроксимируется теоретическим распределением 3-го типа (рисунок 7). параметры которого указаны в таблице 4.

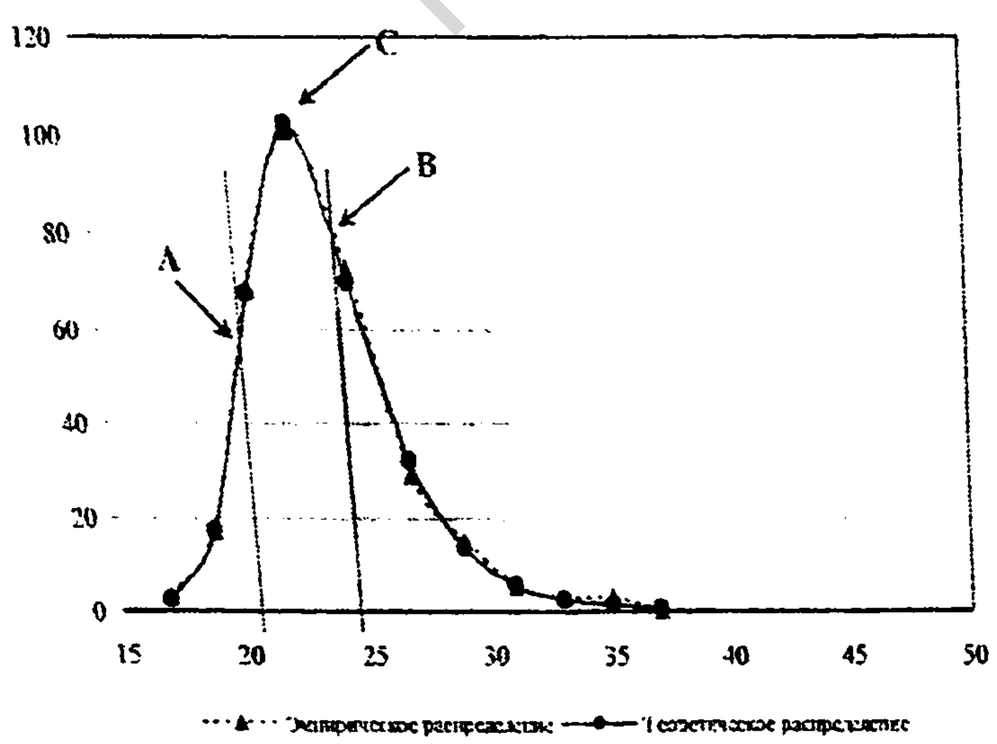

Рисунок 7 – Аппроксимация эмпирического распределения признака «индекс Кетле» теоретическим распределением третьего типа (χ^2=1.143; p=0.9502)

Таблица 4 – Параметры третьего типа распределения признака «индекс Кетле»

Параметр	Значение
αu	–1.96977583 E-20
β	14.60774
γ	19.94334
k = γ / β	2.4346
u	-0.8304599
N	2.01713350 E-26
C (мода). кг/м²	22.89
A. кг/м²	20.88
B. кг/м²	24.87

Графики эмпирического и теоретического распределений признака «индекс Кетле» практически полностью совпадают («сливаются» в одну линию), что свидетельствует о хорошей аппроксимации данным типом обобщенных распределений.

Анализ двумерного распределения «длина тела/масса тела» с учетом новых диапазонов, определенных через точки A, B, C, позволяет разбить всю выборку на 9 классов, которые можно рассматривать как строго обоснованный и в то же время достаточно простой способ антропометрического соматотипирования мужчин в возрасте от 18 до 36 лет (таблица 5).

Таблица 5 – Диапазоны соматотипирования мужчин в возрасте от 18 до 36 лет

Длина тела. м		Масса тела. кг		
		<60,6	60.6–78.9	>78,9
		1	2	3
<1,69	1	*18*	40	8
1,69–1,78	2	14	121	49
>1,78	3	1	34	*31*

Как следует из таблицы 5, предлагаемый способ соматотипирования позволяет выделить три основных соматотипа: лептосомный (1–1) – 18 мужчин (5,7 %), мезосомный (2–2) – 121 мужчина (39,3 %), пикносомный (3–3) – 31 мужчина (9,8 %), а также 6 промежуточных соматотипов (1–2, 1–3, 2–1, 2–3, 3–1, 3–2).

Вычисленные распределения можно представить на плоскости (B; H), где B – показатель асимметрии, H – показатель островершинности аппроксимирующего распределения (рисунок 8). Показатели B и H зависят только от двух параметров формы (k, u), что позволяет изобразить несколько распределений (популяций) на плоскости и наглядно представить их взаимное расположение.

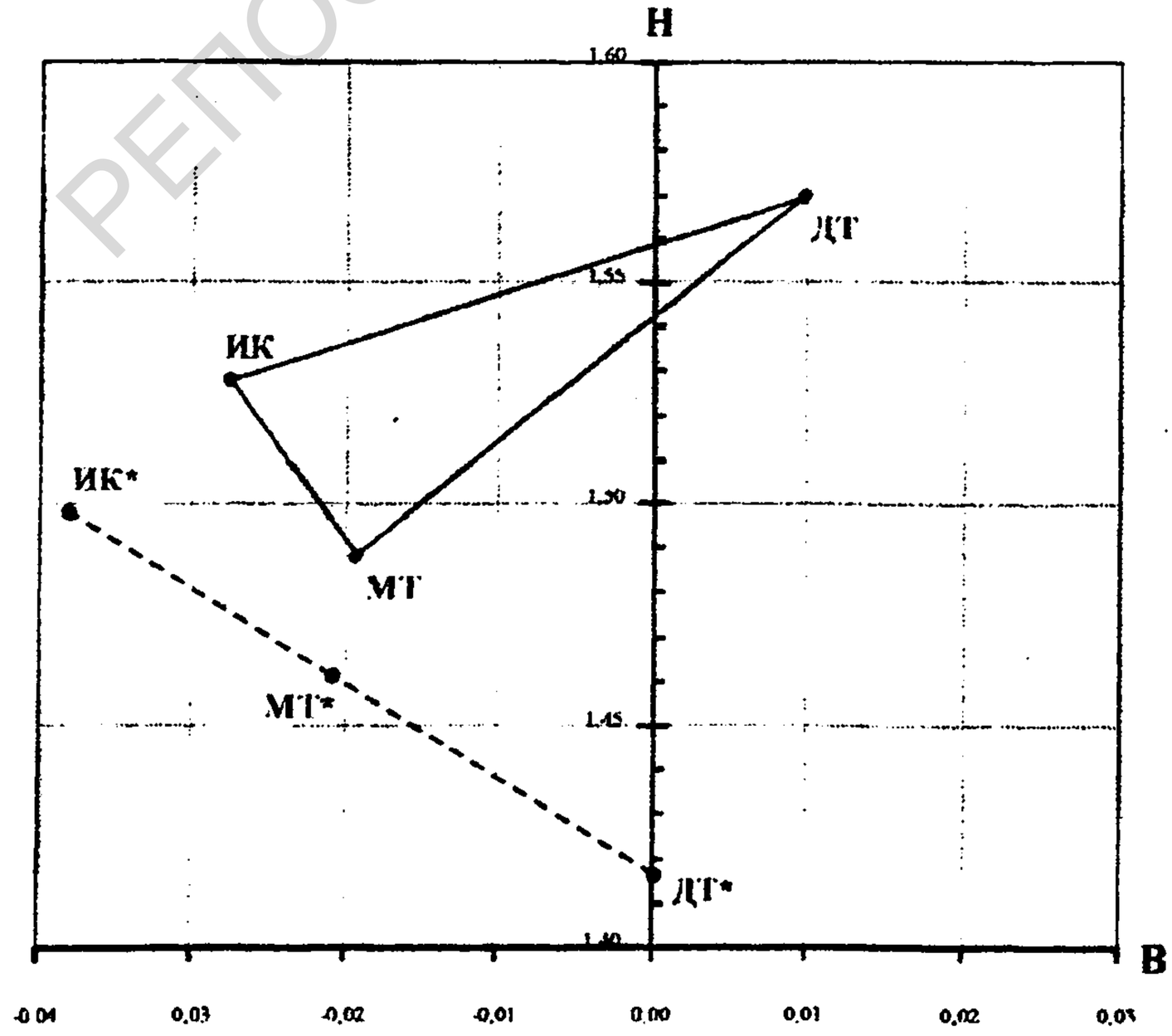

Рисунок 8 – Изображение выравнивающих распределений на плоскости (**B; H**): ДТ, МТ, ИК – мужчины до 36 лет; ДТ*, МТ*, ИК* – мужчины старше 36 лет

На рисунке 8 показано расположение аппроксимирующих распределений признаков «длина тела», «масса тела», «индекс Кетле» в двух выборках мужчин – от 18 до 36 лет (n=316) и от 36 до 61 года (n=266). Визуальное различие аппроксимирующих распределений по указанным признакам более чем очевидно, несмотря на близость средних значений: $ДТ_{316}$=1,734±0,06 м – $ДТ_{266}$=1,696±0,06 м: $МТ_{316}$=71,3±9,6 кг – $МТ_{266}$=72,0±11,6 кг; $ИК_{316}$=23,2±2.9 кг/м^2 – $ИК_{266}$=25,0±3,6 кг/м^2.

Заключение. Проведенное исследование позволяет прийти к утверждению о преимуществах теории обобщенных величин, позволяющей:

1. Вычислить выравнивающее теоретическое распределение путем аппроксимации полученных результатов, исключающей эмпирический подход перебора (подгонки) данных;

2. На основании аппроксимации выделить центральный диапазон, характеризующий популяционную норму распределения, так как большая часть морфофункциональных признаков организма человека имеет высокую вариабельность и не укладывается в закон нормального распределения;

3. Визуализировать аппроксимирующее распределение и сравнить полученные результаты с другими непрерывными теоретическими распределениями.

1. Куршакова. Ю.С. Распределение антропометрических признаков и логарифмическая трансформация / Ю.С. Куршакова // Вопросы антропологии. – 1964. – Вып. 18. – С. 73–79.

2. Лакин. Г.Ф. Биометрия / Г.Ф. Лакин. – М.: Высшая школа. 1990. – 351 с.

3. Нешитой. В.В. Методы статистического анализа на базе обобщенных распределений: учеб.-метод. пособие / В.В. Нешитой. – Минск: Веды. 2001. – 168 с.

4. Нешитой. В.В. // Ученые записки Тартуского университета. – 1987. – Вып. 774. – С. 123–134.

5. Нешитой. В.В. Статистические модели в биологии / В.В. Нешитой // Кибернетика. – 1987. – № 6. – С. 91–96.

6. Нешитой. В.В. Применение обобщенных распределений в системах управления качеством / В.В. Нешитой // Новости стандартизации и сертификации. – 2004. – № 1. – С. 54–58.

7. Нешитой. В.В. Элементы теории обобщенных распределений: монография / В.В. Нешитой. – Минск: РИВШ. 2009. – 204 с.

8. Орлов. А.И. О проверке однородности двух независимых выборок / А.И. Орлов // Заводская лаборатория. – 1991. – Т. 57. – № 7. – С 64–66.

9. Орлов. А.И. Прикладная статистика: учебник / А.И. Орлов. – М.: Экзамен. 2004. – 656 с.

10. Орлов. А.И. О применении статистических методов в медико-биологических исследованиях / А.И. Орлов // Вестник Академии медицинских наук СССР. – 1987. – № 2. – С. 88–94.

11. Реброва. О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ Statistica / О.Ю. Реброва. – М.: МедифСфера. 2003. – 312 с.

12. Салівон. І.І. Фізічны тып беларусаў: Узроставая. тыпалагічная і экалагічная зменлівасть / І.І. Салівон. – Минск: Навука і тэхніка. 1994. – 239 с.

Поступила 07.04.2012