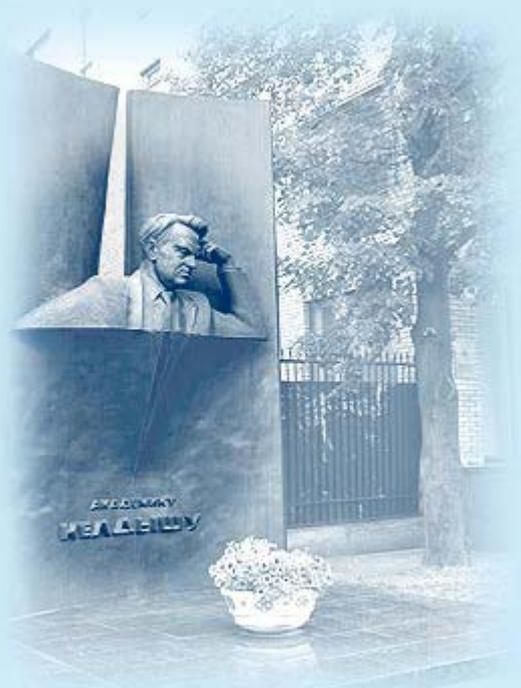




ИПМ им.М.В.Келдыша РАН • [Электронная библиотека](#)

[Препринты ИПМ](#) • [Препринт № 25 за 2020 г.](#)



ISSN 2071-2898 (Print)  
ISSN 2071-2901 (Online)

[Гавриков М.Б.](#), [Кислицын А.А.](#),  
[Орлов Ю.Н.](#), Камбаров А.О.,  
Никитюк Д.Б., Тутельян В.А.

Цифровая  
персонифицированная  
нутрициология: проблемы и  
решения

**Рекомендуемая форма библиографической ссылки:** Цифровая персонифицированная нутрициология: проблемы и решения / М.Б.Гавриков [и др.] // Препринты ИПМ им. М.В.Келдыша. 2020. № 25. 36 с. <http://doi.org/10.20948/prepr-2020-25>  
URL: <http://library.keldysh.ru/preprint.asp?id=2020-25>

**Ордена Ленина  
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ  
имени М.В.Келдыша  
Российской академии наук**

**М.Б. Гавриков, А.А. Кислицын, Ю.Н. Орлов,  
А.О. Камбаров, Д.Б. Никитюк, В.А. Тутельян**

**Цифровая персонифицированная  
нутрициология:  
проблемы и решения**

**Москва — 2020**

**Гавриков М.Б., Кислицын А.А., Орлов Ю.Н., Камбаров А.О.,  
Никитюк Д.Б., Тутельян В.А.**

Цифровая персонифицированная нутрициология: проблемы и решения

Формулируются основные положения новой прикладной науки – цифровой нутрициологии. Обсуждаются методы исследования и модели, которые могут быть использованы в качестве основы для вычислительных алгоритмов в этой области. Рассматривается связь между питанием и продолжительностью жизни. Изучаются вопросы точности исходных данных, эмпирических зависимостей и вычислительных процедур. Анализируется математическая модель расчета состава и величины порций оптимального меню в зависимости от антропометрических характеристик пользователя.

**Ключевые слова:** цифровая персонифицированная нутрициология, математическое моделирование, оптимальное питание

**Gavrikov M.B., Kislitsyn A.A., Orlov Yu.N., Kambarov A.O., Nikitjuk D.B.,  
Tutelyan V.A.**

Numerical personal threpsology: the problems and solutions

The main provisions of the new applied science – digital nutrition – are formulated. The research methods and models that can be used as a basis for computational algorithms in this area are discussed. The relationship between nutrition and life expectancy is considered. We study the accuracy of source data, empirical dependencies, and computational procedures. The mathematical model for calculating the composition and values of portions of the optimal menu is analyzed, depending on the anthropometric characteristics of the user.

**Keywords:** numerical personal threpsology, mathematical modeling, optimal nutrition

## Содержание

1. Введение.....	3
2. Классификация параметров пользователей.....	7
3. Структура потребления продуктов питания и нутриентов.....	15
4. Математические модели в задачах нутрициологии.....	18
5. Базовые меню.....	28
6. Расчет величин порций блюд.....	31
7. Заключение.....	34
Список источников.....	35

## 1. Введение

Данная работа направлена на систематизацию подходов, методов и экспериментальных данных, которые необходимо использовать для решения задач, возникающих в области цифровой нутрициологии.

Нутрициология – это медицинская специальность, изучающая фундаментальные механизмы ассимиляции пищевых и биологически активных веществ в норме и патологии, состояние питания различных групп детского и взрослого населения, качество и безопасность пищевой продукции, включая новые источники пищи и продукции, полученной с использованием современных био- и нанотехнологий, синтетической биологии, раннюю диагностику, лечение и профилактику алиментарно-зависимых состояний и заболеваний, направленные на увеличение продолжительности и повышение качества жизни детского и взрослого населения.

Как следствие, эта область знаний содержит большое количество информации числового характера, такой, например, как процентный химический состав пищевых продуктов или структура человеческой популяции по полу, возрасту, массе тела и иным измеримым параметрам, а также и качественную информацию рангового типа, относящуюся к индивидуальным особенностям потребления продуктов питания, связанным с культурными, социально-экономическими и иными различиями. Разнородность данных числового и рангового типов затрудняет использование формализованных моделей питания отдельного индивидуума в виде математических алгоритмов и требует применения статистических методов. Цифровизация как раз и направлена на алгоритмическое соединение условно «достоверного» знания, реализованного в виде детерминистических моделей и формул с привязкой к базам данных в электронном формате, и условно «усредненного» знания, полученного статистическим путем без выявления прямых причинно-следственных связей. Строго говоря, «достоверное» знание на самом деле тоже имеет статистический характер, просто неопределенность в получении выводов в рамках таких моделей меньше, чем практически требуемая точность. Иными словами, дисперсия функции распределения «достоверных» параметров мала, а «недостоверные» данные требуют наилучшей кластеризации для уменьшения дисперсии внутри класса. В этом смысле задача состоит в построении совместных распределений для более и менее достоверно определяемых параметров так, чтобы суммарная ошибка была бы минимальна.

Следует подчеркнуть, что статистические методы эффективны именно при адекватной кластеризации параметров, что не всегда может быть выполнено чисто алгоритмическим путем, поскольку границы между классами очерчены нечетко. Вопросы группировки требуют привлечения экспертного знания, что также связано с его цифровой формализацией.

Например, в настоящей работе мы обсуждаем вопросы питания взрослого населения, т.е. людей старше 18 лет. Однако возраст есть косвенный параметр,

в действительности вопрос связан не с ним, а с формированием устойчивого уровня основного обмена веществ в организме, который индивидуально может быть достигнут как в 16 лет, так и к 20 годам быть еще не сформировавшимся. В этой связи любая формально проведенная граница условна и будет приводить как к ошибочному включению в статистику, так и к ошибочному исключению из нее. Люди старше 60 лет в большинстве своем отличаются, конечно, от тех, кто моложе 18 лет, но вот провести границу между теми, кому 25 лет, и теми, кому 26 лет, затруднительно, если не основываться непосредственно на метрических записях. Следовательно, людей надо определенным образом сгруппировать в классы с близкими значениями косвенных параметров, причем таким образом, чтобы разные классы существенно различались бы по значениям прямых параметров, что само по себе не есть простая задача. Кроме того, формализация различия между классами требует введения соответствующего пространства классов и функции расстояния в нем: можно сравнивать средние значения группы параметров из разных классов, функции распределения параметров в разных классах в норме суммируемых или непрерывных функций и т.д. Также надо провести кластеризацию расстояний и между классами, чтобы установить, какие различия между ними считать существенными, а какие – нет. Тем самым проблема оптимальной классификации переносится на следующий уровень, где также надо проводить соответствующий статистический анализ. Такое погружение могло бы быть бесконечным, если бы количество исходных данных было бесконечным. Но в силу конечности набора данных описанный процесс остановится тогда, когда статистическая неопределенность сравняется с расстоянием между классами, т.е. на практике будет достигнута сходимость, понимаемая как наилучшая эмпирическая точность.

Мы подробно остановились на этом техническом вопросе, чтобы показать, что решение ряда проблем, кажущихся интуитивно ясными, лежит в весьма сложной области математической статистики.

В то же время собственно моделей в области питания (или диет, или, более общо, паттернов меню) существует довольно много. Определить без участия специалиста, какой режим питания наилучшим образом отвечает конкретному индивидууму, часто бывает затруднительно. Трудность усугубляется также и тем, что понятие «наилучшим образом» не вполне конкретно и допускает разные целевые функции. Кроме того, разными способами можно достичь одного и того же, так что решение может быть неоднозначным. Цифровизация направлена в том числе и на то, чтобы, не заменяя врача-диетолога, очертить наиболее вероятную область возможностей в персонифицированном выборе режима питания.

Эта ситуация аналогична различию между просто «моделированием» и «математическим моделированием». Если первое представляет собой конкретную реализацию определенной инженерной идеи, которую без описания алгоритма практически невозможно повторить другому человеку, то второе содержит описание свойств модели на уровне математических теорем, что позволяет автоматически воспроизводить тот или иной класс решений.

Таким образом, целью цифровизации знания, которое отчасти «гуманитарно», т.е. не имеет вида числовой функциональной связи, является представление этого знания в форме, допускающей алгоритмические действия с прогнозируемой точностью результата. В частности, необходимость использования автоматической машинной классификации в области нутрициологии обусловлена не только большим объемом накопленной информации о связи питания с продолжительностью жизни и здоровьем, требующим компьютерной обработки больших массивов данных, но и социально-экономическим аспектом планирования производства продуктов питания, что также требует использования вычислительных средств в большом объеме.

В соответствии с принципами цифровизации большой совокупности данных, часть которых имеет не числовые, а ранговые или иные характеристики, применительно к проблемам нутрициологии вырисовывается следующая концептуальная последовательность действий по разработке структуры вычислительной платформы для обработки данных по продуктам питания и блюдам и для создания персонализированных рекомендаций по оптимальному питанию.

1. Необходимо сформировать структуру массива данных о физиологических потребностях человека в пищевых и биологически активных веществах и энергии, а также данных о химическом составе основных пищевых продуктов. Этот массив будет иметь два аспекта своей реализации: априорный и персональный. Априорный вариант представляет совокупность данных по сбалансированному потреблению для здорового человека в зависимости от его достоверно определяемых основных параметров (пол, возраст, рост, масса тела, физическая активность особенности генома). Эти данные выражают современные представления диетологии. Персональный вариант также является усредненным, но получается он в результате оптимальной кластеризации параметров, уменьшающей дисперсию рекомендаций по оптимальному питанию. Этот второй вариант получается в результате статистической обработки данных по людям (далее условно называемых «пользователями»), чьи детализированные параметры попадают в одинаковые классы. Сравнение априорного и персонального вариантов позволит сделать предварительную оценку оптимальности питания конкретного пользователя, на основе которой можно будет строить уточняющие персонализированные рекомендации.

2. Используя структуру данных по продуктам питания и модельные представления об энергозатратах, необходимо создать вычислительную программу для выработки рекомендаций по оптимальному питанию. Входные и выходные данные для работы этой программы определяются согласно базе данных продуктов и блюд (далее БДПБ). Основными входными параметрами являются биометрические данные пациента, его национальные особенности, место рождения и место постоянного проживания, пищевые предпочтения, режим питания, алиментарно-зависимые заболевания. Выходными параметрами являются продукты питания в виде набора продуктовой корзины и

в виде ориентировочного меню блюд с суточной периодичностью на заданный временной горизонт планирования.

3. Следует разработать статистические модели по особенностям питания, связанным с региональным распределением населения РФ. Для этой цели необходимо использовать медицинские рекомендации, данные оценочных расчетов изменения региональной структуры населения по полу и возрасту, экологической обстановки, национальных особенностей для прогнозирования производства необходимых продуктов для рационального питания по регионам.

4. На основе обработки неперсонифицированных данных по диетологии необходимо создать программу по разработке персонифицированных рекомендаций по оптимальному питанию на заданный горизонт планирования. Как уже говорилось выше, этот аспект статистического анализа включает в себя элементы машинного обучения применительно к пользователям, которые обладают сходными параметрами.

Таким образом, создание программного продукта для цифровой нутрициологии предполагает решение следующих практических задач.

1. Проведение цифровой трансформации данных о физиологических суточных потребностях человека в пищевых и биологически активных веществах и энергии. Обработка этих данных должна быть представлена в виде совокупности вычислительных алгоритмов, сопоставляющих многомерному массиву параметров пользователя плотность распределения рекомендуемой величины потребления каждого из нутриентов с указанием среднего значения и допустимого стандартного отклонения.

2. Создание иерархической БДПБ по химическому составу основных пищевых продуктов и блюд с указанием точности приводимых данных. Иерархия относится к режимам питания (количество раз в день), а также к национальным и культурным особенностям пользователя.

3. Разработка статистических моделей по особенностям питания, т.е. создание базовых меню для пользователей и типовых перечней продуктов, попадающих в определенные категории по своим параметрам. Для этой цели используются усредненные медицинские рекомендации, количественный прогноз возрастной структуры населения, экологической обстановки.

4. Создание программы по разработке персонифицированных рекомендаций по оптимальному питанию на основе машинной классификации результатов обработки неперсонифицированных данных в этой области.

5. Создание модели прогнозирования регионального производства необходимых продуктов для рационального питания.

Следует сказать, что в настоящее время в России отсутствует комплексная вычислительная программа, охватывающая не только персонифицированные, но и укрупненные региональные аспекты нутрициологии. Не разработана и методика построения соответствующих алгоритмов с использованием как детерминистических моделей типа энергетического баланса, так и статистических моделей, использующих результаты обработки большого объема персонифицированных данных и их классификации. Не сформулированы в виде математического критерия понятия оптимальности

структуры питания, а существующие многочисленные индексы питания (здорового, разнообразного и т.п.) не находятся во взаимно однозначном соответствии между собой.

В настоящей работе формулируется подход к решению перечисленных задач и описываются проблемные места, требующие новых идей как в части сбора и кластеризации исходной информации, так и в области разработки математических моделей и вычислительных алгоритмов.

## **2. Классификация параметров пользователей**

### *2.1. Основные биометрические параметры*

С каждым пользователем связывается основное 5-мерное пространство параметров. Эти параметры следующие: 1 – пол, 2 – возраст, 3 – масса тела, 4 – рост (длина тела), 5 – уровень физической активности.

Дополнительные биометрические параметры, влияющие на особенности питания, относятся к более сложно определяемым величинам – это тип метаболизма, величина подкожного жира, тип скелета.

Также необходимо ввести классы пользователей в соответствии с их культурными особенностями и социально-экономическим положением.

На структуру питания влияет экологическая обстановка в области проживания и другие внешние факторы (вид трудовой деятельности, наличие домашнего участка).

Кроме того, в перечень параметров входят алиментарные заболевания, которые могут быть у пользователя и с которыми связаны специфические продукты питания или их изъятия.

В настоящей работе мы рассматриваем только основные параметры, которые имеют следующую внутреннюю классификацию для здоровых людей.

Два класса полов: 1 – мужской пол, 2 – женский пол.

Четыре возрастные группы населения: 1 – [18-29] лет, 2 – [30-39] лет, 3 – [40-59] лет, 4 – лица пожилого возраста (старше 60 лет).

Девять классов по массе тела: от 50 до 90 кг для мужчин и от 40 до 80 кг для женщин с шагом в 5 кг.

Пять ранговых параметров для классификации физической активности. Согласно методическим рекомендациям Минздрава [1, 2], для взрослого населения вводится: 1 – очень низкая активность энерготрат, ранговый индекс  $k = 1,4$ ; 2 – низкая активность, ранговый индекс  $k = 1,6$ ; 3 – средняя активность, ранговый индекс  $k = 1,9$ ; 4 – высокая активность, ранговый индекс  $k = 2,2$ ; 5 – очень высокая активность, ранговый индекс  $k = 2,5$ .

Таким образом, в представленной укрупненной классификации размерность вектора параметров пользователя, равная произведению количества независимых классов, составляет  $n = 360$  без учета дополнительных параметров, перечисленных выше. Предполагая, что каждая компонента этого вектора определяется на основе практических измерений, т.е. имеет

статистический характер, для оценки относительной точности  $\varepsilon$  получаемых данных по выборке конечного объема  $N$  используем формулу [3]

$$N = -\frac{\pi n}{2\varepsilon^2} \ln(1 - (1 - \varepsilon)^2). \quad (2.1)$$

Из формулы (2.1) следует, что для получения статистики с точностью  $\varepsilon = 0,1$  достаточно получить данные приблизительно от  $N = 94$  тыс. респондентов, для точности  $\varepsilon = 0,05$  потребуется уже примерно 526 тыс. данных, а для точности  $\varepsilon = 0,01$  необходимо собрать более 22 млн персональных данных. Более того, статистика по алиментарным заболеваниям заведомо недостаточна для получения желаемой точности. Поскольку получение точности даже на уровне 10 % представляется пока не вполне достижимой целью, наряду со статистическими исследованиями необходимо использовать различные теоретические модели о связи рассматриваемых параметров с калорийностью питания и с потреблением эссенциальных химических веществ и соединений.

## 2.2. Половозрастная структура населения

Несмотря на то что возрастных групп введено всего четыре, для анализа динамики состава этих групп по годам и изучения связи между типом питания и продолжительностью жизни требуется более детальная статистика, с шагом в один год. Вообще все модели, имеющие отношение к статистике населения, основываются на демографических показателях. Эти показатели следующие: половозрастной состав населения  $N^{F,M}(a,t)$  как число (в тысячах) женщин (индекс  $F$ ) и мужчин (индекс  $M$ ) возраста  $a$  лет в год  $t$ ; повозрастной коэффициент рождаемости  $B^{F,M}(a,t)$  как число родившихся девочек и мальчиков на 1 тысячу женщин возраста  $a$  лет в год  $t$ ; повозрастной коэффициент смертности  $q^{F,M}(a,t)$  как число умерших женщин и мужчин на 1 тысячу соответствующих групп возраста  $a$  лет в год  $t$ .

Хотя повозрастные коэффициенты рождаемости и смертности зависят от текущего года, т.е. они не постоянны, достоверных моделей их зависимости от времени в настоящее время нет. Поэтому для прогноза динамики численности населения используется сценарный подход, где выбирается базовое (как в последний год), повышенное на 10-15 % или пониженное значения коэффициентов. Мы будем опираться на базовый сценарий, взяв за основу данные Росстата за 2017 г. и считая коэффициенты  $B^{F,M}(a,t)$  и  $q^{F,M}(a,t)$  зависящими только от возраста человека. Эти данные приведены соответственно на рис. 1-3 (см. также [4]).

Пусть в год  $t$  население характеризовалось численностью  $N^{F,M}(a,t)$ . Тогда в следующий год  $t+1$  численность населения возраста  $a$  будет равна

$$N^{F,M}(a,t+1) = N^{F,M}(a-1,t) \cdot (1 - q^{F,M}(a-1)). \quad (2.2)$$

В соответствии с формулой (2.2) численность возрастных классов, введенных в п. 2.1, меняется со временем в зависимости от исходной

возрастной структуры населения. Например, численность первого класса в год  $t$  равна

$$N_1^{F,M}(t) = \sum_{a=18}^{29} N^{F,M}(a,t). \quad (2.3)$$

Тогда в год  $t+1$  она составит

$$N_1^{F,M}(t+1) = \sum_{a=17}^{28} N^{F,M}(a,t) \cdot (1 - q^{F,M}(a)). \quad (2.4)$$

Аналогичные формулы выписываются и для численностей групп других классов. Следовательно, групповая численность меняется со временем. Поскольку же калорийность питания зависит в том числе и от возраста (см. далее раздел 3), то структура оптимальной продуктовой корзины, вообще говоря, меняется от года к году просто в силу естественных демографических причин, не считая социально-экономических.

Однако следует подчеркнуть, что данные Госкомстата о структуре питания домашних хозяйств содержат информацию о среднедушевом официальном доходе семьи, но не отражают биометрической информации, что препятствует разработке корректной статистической модели фактического потребления продуктов питания в России с точки зрения оптимальности в понимании нутрициологии. Иными словами, разрабатываемую модель трудно будет верифицировать. Можно лишь предположить независимость вводимых эмпирических распределений.

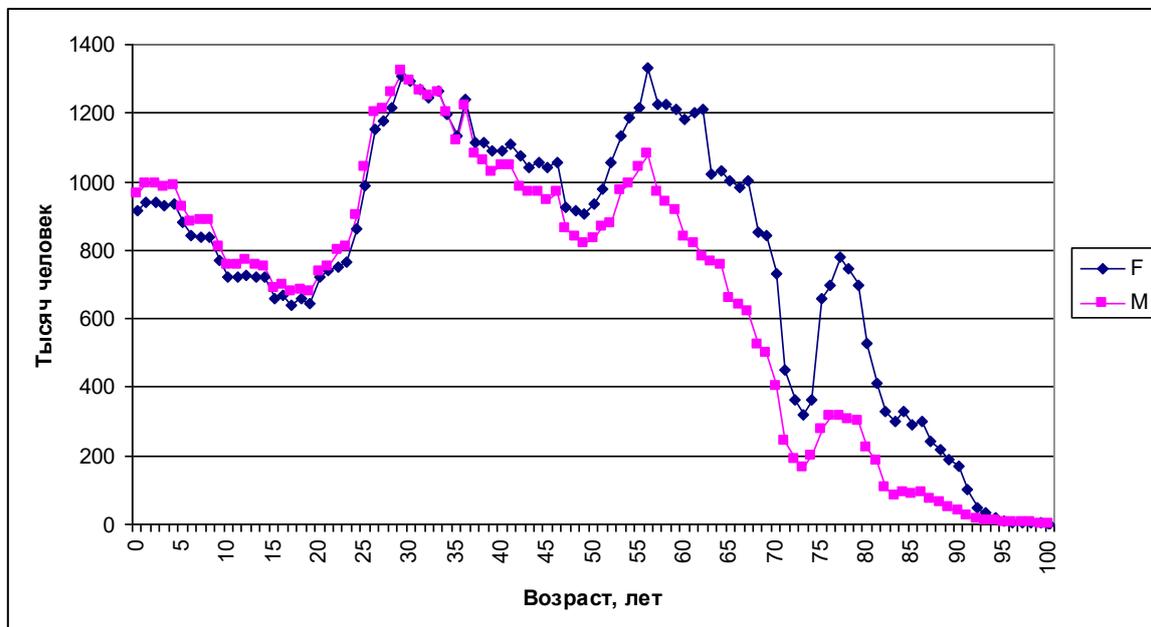


Рис. 1 – Распределение населения РФ по полу и возрасту на 2017 г.

Как можно видеть, население России в силу ряда внешних и внутренних причин имеет пики рождаемости (демографические волны), приходящиеся на 1988, 1961 и 1940 годы, а также провалы в 2002, 1968 и 1944 годах. Однако изрезанность графика приводит к тому, что численность возрастных групп

согласно (2.3) значительно меняется с течением времени, что заметно влияет на рекомендуемые объемы производства и потребления продуктов питания в масштабе страны.

На рис. 2 приведено количество рождений  $B(a)$  мальчиков и девочек на 1 тыс. женщин соответствующего возраста  $a$  по данным [4] на 2016 г. Рождения отдельно мальчиков и отдельно девочек определяются из этих данных путем введения коэффициентов  $B^M(a) = 0,512B(a)$ ,  $B^F(a) = 0,488B(a)$ .

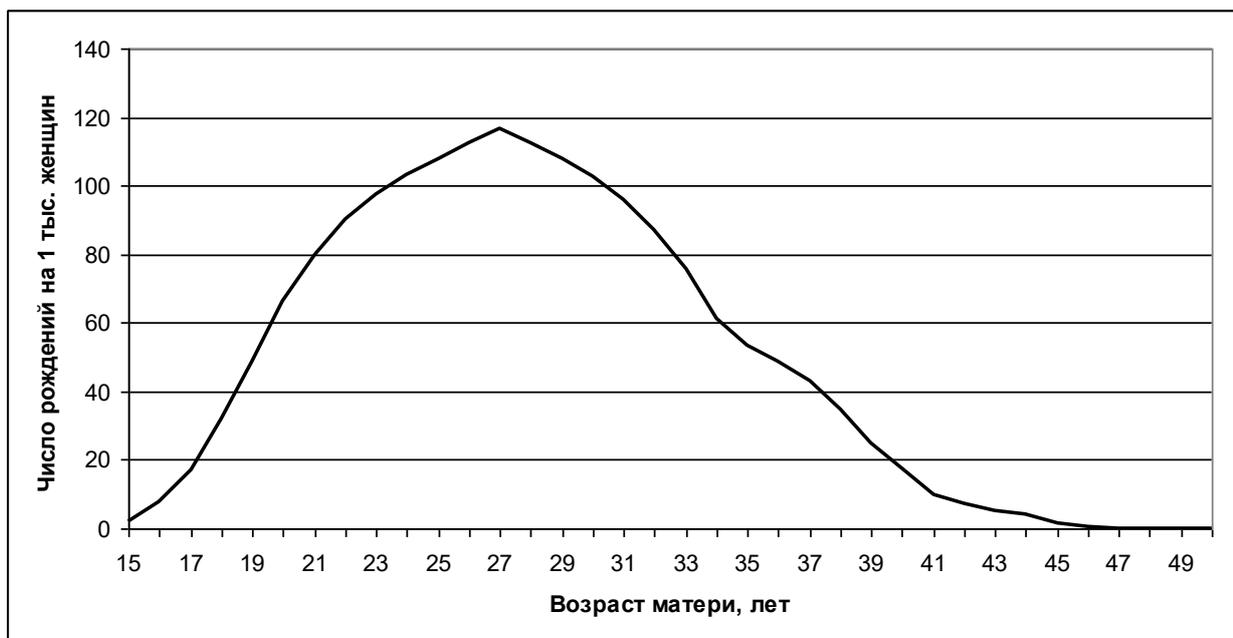


Рис. 2 – Повозрастные коэффициенты рождаемости в РФ на 2016 г.

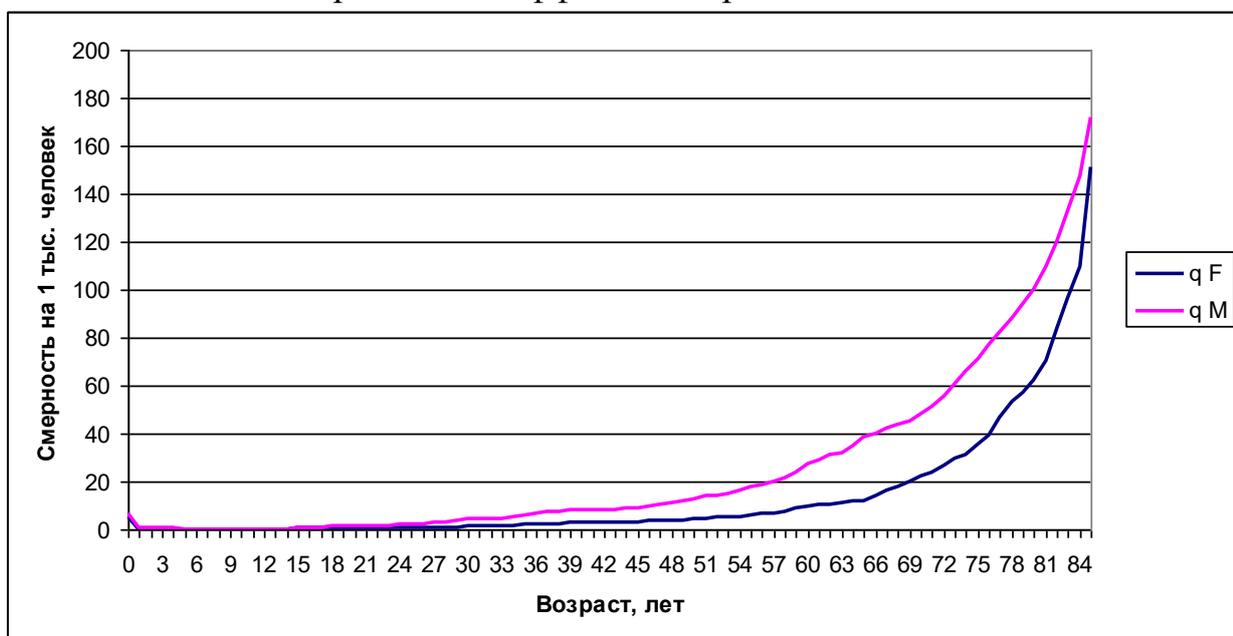


Рис. 3 – Повозрастные коэффициенты смертности в РФ на 2016 г.

Относительное изменение повозрастных коэффициентов рождаемости и смертности от года к году различно в разных возрастных группах. В средних возрастных группах оно составляет величину порядка 7 %, в младших возрастах оно увеличивается до 16 %, а в старших уменьшается до 5 %. Это

позволяет оценить точность прогноза численности возрастных групп. Пусть  $\delta_k^*$  есть максимальное относительное изменение коэффициента смертности в  $k$ -ой группе. Определим тогда относительную неточность прогноза численности  $k$ -ой возрастной группы на следующий год. В наблюдаемый год  $t$  ее численность определяется формулой (для краткости индексы  $F, M$  временно опущены):

$$N_k(t) = \sum_{a=a_k}^{b_k} N(a,t),$$

где  $a_k, b_k$  – границы группы в соответствии с введенными классами. В следующий год она, предположительно, будет равна

$$\tilde{N}_k(t+1) = \sum_{a=a_k-1}^{b_k-1} N(a,t) \cdot (1 - q(a)),$$

тогда как по факту окажется, что

$$N_k(t+1) = \sum_{a=a_k-1}^{b_k-1} N(a,t) \cdot (1 - (1 + \delta_k(a)) \cdot q(a)).$$

Поскольку

$$\left| \tilde{N}_k(t+1) - N_k(t+1) \right| = \sum_{a=a_k-1}^{b_k-1} N(a,t) \cdot q(a) \cdot \delta_k(a) \leq \delta_k^* \sum_{a=a_k-1}^{b_k-1} N(a,t) \cdot q(a),$$

то максимальная относительная ошибка такого прогноза в численности группы в линейном приближении по коэффициенту смертности не превосходит произведения максимального относительного отклонения  $\delta_k^*$  на групповой коэффициент смертности в группе, сдвинутой вниз на 1 год по сравнению с исходной:

$$\frac{\left| \tilde{N}_k(t+1) - N_k(t+1) \right|}{N_k(t+1)} = \frac{\left| \tilde{N}_k(t+1) - N_k(t+1) \right|}{N_k^-(t)} + o(q_k^*) \leq \delta_k^* q_k^* + o(q_k^*) \quad (2.5)$$

$$q_k^* = \frac{1}{N_k^-(t)} \sum_{a=a_k-1}^{b_k-1} N(a,t) \cdot q(a), \quad N_k^-(t) = \sum_{a=a_k-1}^{b_k-1} N(a,t).$$

Формула (2.5) показывает, что предположительное изменение продолжительности жизни, связанное с уменьшением коэффициента смертности путем перехода на оптимальное питание в некоторой возрастной группе, может быть доказательно проверено, только если относительное изменение численности возрастных групп превосходит величину  $\delta_k^* q_k^*$ . Для групп, введенных в п. 2.1, вычисленные по данным за 2016 г. по формуле (2.5) групповые коэффициенты смертности следующие (таблица 1):

Табл. 1 – Групповые коэффициенты смертности и численность групп, млн чел.

Возраст		18-29	30-39	40-59	> 60
<i>F</i>	<i>q*</i>	0,66	1,75	4,37	29,00
	<i>N</i>	11,0	12,0	21,7	19,7
<i>M</i>	<i>q*</i>	1,97	5,23	12,53	48,75
	<i>N</i>	11,4	11,8	18,9	10,8

### 2.3. Распределение населения по росту и массе тела

Масса тела человека – параметр, который меняется в пределах нескольких процентов в течение суток и имеет некоторые тенденции к увеличению среднегодовых значений с возрастом. Поэтому данные о распределении населения по массе тела, во-первых, достаточно условны и, во-вторых, косвенные. Сравнительно постоянной величиной является рост человека после 18 лет. Однако надо учитывать, что в пожилом возрасте (после 50 лет) рост уменьшается в среднем на 1 см за 10 лет.

Распределение населения по росту используется для того, чтобы на основе выборочного анализа оценить параметр, называемый индексом массы тела  $I$  (ИМТ), определяемый как отношение массы  $m$  [кг] к квадрату роста  $h^2$  [м<sup>2</sup>]. Согласно рекомендациям ВОЗ [5], если  $I \leq 16$ , то имеется выраженная худоба; если  $16 < I < 16,99$  – умеренная худоба,  $17 < I < 18,49$  – легкая худоба; нормальные пределы соответствуют  $18,5 < I < 24,99$ ; избыточная масса тела  $I \geq 25,0$ ; предожирение –  $25,0 < I < 29,99$ ; если  $30,0 < I < 34,99$ , то классифицируется ожирение I степени; если  $35,0 < I < 39,99$ , то ожирение II степени, если  $I \geq 40,0$ , то ожирение III степени. В то же время следует учитывать, что данная классификация отвечает среднему уровню физической активности, тогда как для высокого уровня (тяжелой физической работы или для спортсменов) этот индекс может сдвигаться вправо.

Согласно исследованиям [6], для возрастных групп ИМТ приближенно моделируется нормальным (гауссовым) распределением, параметры которого приведены ниже в таблице 2 и зависят от возраста  $a$ .

Плотность нормального распределения в терминах данных таблицы 2 дается формулой:

$$f(I, a) = \frac{1}{\sqrt{2\pi}\sigma_I(a)} \exp\left(-\frac{(I - \langle I \rangle(a))^2}{2\sigma_I^2(a)}\right). \quad (2.6)$$

Заметим, что, поскольку (2.6) есть приближенное распределение реальных данных, то на практике оно перенормируется на пределы  $I_{\min} \div I_{\max}$  возможных значений ИМТ.

Табл. 2 – Параметры распределения ИМТ по возрастным группам

Возраст, $a$ лет	Пол	$\mu = \langle I \rangle(a)$	$\sigma_I(a)$	$I_{\min}(a)$	$I_{\max}(a)$
18-29	$F$	22,20	3,94	14,20	43,82
	$M$	23,40	3,47	14,45	41,91
30-44	$F$	25,76	5,23	16,22	51,26
	$M$	25,46	3,75	15,78	43,21
45-59	$F$	28,90	5,51	14,87	49,67
	$M$	26,33	4,35	14,18	47,75
60-69	$F$	29,14	5,13	16,41	50,19
	$M$	26,51	4,34	17,31	39,56
70-79	$F$	28,89	5,87	15,06	62,65
	$M$	26,43	4,10	17,72	41,04
$\geq 80$	$F$	27,25	5,74	15,62	46,87
	$M$	26,48	4,02	16,73	36,20

Распределение же людей по росту  $h$  считается не зависящим от возраста, и в целом по популяции людей старше 18 лет также моделируется гауссовым распределением с плотностью

$$f(h) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(h-\mu)^2}{2\sigma^2}\right), \quad (2.7)$$

где параметры  $\mu$  и  $\sigma$  для женщин и мужчин в современной России равны соответственно  $\mu^F = 159,0$ ,  $\mu^M = 172,5$ ,  $\sigma^F = 6,3$ ,  $\sigma^M = 7,0$ .

Плотность распределения квадрата роста  $y = h^2$  имеет тогда плотность

$$\varphi(y) = \frac{1}{2\sqrt{2\pi}\sigma} \exp\left(-\frac{(\sqrt{y}-\mu)^2}{2\sigma^2}\right). \quad (2.8)$$

Исходя из того, что плотность распределения величины  $I$  известна и дается формулой (2.6), а масса есть результат произведения двух случайных величин  $m = Iy$ , то плотность совместного распределения населения по массе тела и возрасту  $\psi(m, a)$  дается формулой

$$\psi(m, a) = \int_{-\infty}^{\infty} \varphi(y) f(m/y, a) \frac{dy}{|y|} \quad (2.9)$$

и может быть определена численным интегрированием (суммированием соответствующей дискретной аппроксимации) с учетом (2.7), (2.8) и таблицы 2. В результате получаются кластерные распределения  $\psi(i, j)$  для женщин и мужчин, находящихся в  $i$ -ой группе по массе тела и в  $j$ -ой группе по возрасту (см. п. 2.1).

Распределение пятого параметра – уровня физической активности в зависимости от остальных трех (пол, возраст, масса тела, рост) – известно

гораздо менее точно. Данными могут служить выборочные отчеты Министерства спорта и Минздрава. Согласно этим отчетам, в целом можно считать, что активный образ жизни (с коэффициентом выше среднего) ведут примерно 60 млн человек. Из них 60 % - мужчины, 40 % - женщины, возрастной диапазон на 70 % принадлежит первой группе, 20 % - второй и 10 % - третьей.

К профессиональному спорту (5-я категория активности) относится около 3 млн человек, из них 60 % - мужчины, 95 % которых (как и женщин) принадлежит первой возрастной группе.

В очень низкую группу активности следует отнести людей, страдающих ожирением и некоторых слишком худых – к этому классу относится примерно 2 млн человек. Кроме того, в группу очень низкой активности надо отнести примерно треть людей старше 80 лет, т.е. еще 1 млн.

Просто малоподвижный образ жизни ведут около 20 млн человек.

Остальные 30 млн человек имеют средний уровень активности, который распределен предположительно равномерно по полу и возрасту.

В итоге имеем следующую таблицу 3, в которой приведены доли населения с определенным типом физической активности, нормированные на численность соответствующего возрастного диапазона (см. таблицу 1). Комбинируя эти оценки с данными таблицы 1, получаем, что средний коэффициент физической активности как для мужчин, так и для женщин и для всего населения в целом равен 1,8, т.е. население в целом принадлежит к средней группе активности.

Табл. 3 – Распределение населения по уровню физической активности

Возраст, $a$ лет	Пол	$k = 1,4$	$k = 1,6$	$k = 1,9$	$k = 2,2$	$k = 2,5$
18-29	$F$	0,01	0,05	0,84	0,10	-
	$M$	0,01	0,05	0,57	0,22	0,15
30-39	$F$	0,02	0,05	0,92	0,01	-
	$M$	0,02	0,05	0,89	0,03	0,01
40-59	$F$	0,02	0,10	0,88	0	-
	$M$	0,02	0,10	0,87	0,01	0
$\geq 60$	$F$	0,03	0,60	0,37	0	-
	$M$	0,03	0,60	0,36	0,01	0

Подчеркнем, что анализ точности данных и результатов измерений, а также модельных предположений является важнейшим этапом создания вычислительных алгоритмов в области нутрициологии. Поэтому мы достаточно много внимания уделили анализу исходных данных, касающихся собственно потребителей. Обратимся теперь к статистике потребления ими продуктов питания и содержащихся в этих продуктах нутриентов.

### 3. Структура потребления продуктов питания и нутриентов

#### 3.1. Балансируемые нутриенты и калорийность питания

Согласно [1, 2], нормы физиологических потребностей (НФП) в энергии и пищевых веществах – это усредненная величина необходимого поступления пищевых и биологически активных веществ, обеспечивающая оптимальную реализацию физиолого-биохимических процессов, закрепленных в генотипе человека. Эти нормы являются научной базой при планировании объемов производства основного продовольственного сырья и пищевых продуктов в РФ с учетом изменения социально-экономической ситуации и демографического состава населения для обоснования оптимального развития агрокомплекса и обеспечения продовольственной безопасности страны. Кроме того, на их основе организуется планирование питания в рабочих коллективах и лечебно-профилактических учреждениях, разрабатываются рекомендации по питанию для различных групп населения, обосновывается состав специализированных и обогащенных пищевых продуктов. Также нормы служат критерием оценки фактического питания на индивидуальном и популяционном уровнях.

Основные нутриенты, которые требуются организму, следующие.

Макронутриенты: 1 – белок, 2 – жир, 3 – углевод.

Микронутриенты: 4 – витамин С, 5 – витамин А, 6 – К (калий), 7 – Са (кальций), 8 – Р (фосфор), 9 – Mg (магний), 10 – Na (натрий).

Микронутриенты малых концентраций: 11 – витамин В1, 12 – витамин В2, 13 – витамин РР, 14 – каротин, 15 – Fe (железо).

Макронутриенты обеспечивают организм энергией. Калорийность белка составляет  $e_1 = 4$  ккал/г, жира  $e_2 = 9$  ккал/г, углеводов  $e_3 = 4$  ккал/г. Эти данные используются для оценки калорийности питания в соответствии с тем, сколько макронутриентов содержат те или иные продукты и блюда из них.

Следует отметить, что существует довольно много линейных регрессионных аппроксимаций эмпирических данных потребности в энергии в зависимости от биометрических показателей (формулы Миффлина–Сан-Джеора, Харриса–Бенедикта, Кэтча–МакАрдила, рекомендации ВОЗ и др.). Поскольку такие модели не имеют корректной верификации применительно к населению РФ, то в нашем анализе они использоваться не будут.

Нормативные потребности в энергии  $E_{ij}^{F,M}$  для женщин и мужчин в  $i$ -ой группе по массе тела и  $j$ -ой группе по возрасту согласно рекомендациям [1, 2] в соответствии с изменением величины основного обмена (далее ВОО) у человека приведены в таблицах 4 и 5.

В целом можно сказать, что ВОО снижается с возрастом и увеличивается с увеличением массы тела. Эмпирическая кривая приблизительно постоянных энергозатрат отвечает массе тела, увеличивающейся с возрастом как  $m \propto a^{1/7}$ .

Табл. 4 – Нормативные потребности в энергии  $E_{ij}^F$  для женщин, ккал/сут

	$j=1$	$j=2$	$j=3$	$j=4$
$i=1$	1080	1050	1020	960
$i=2$	1150	1120	1080	1030
$i=3$	1230	1190	1160	1100
$i=4$	1300	1260	1220	1160
$i=5$	1380	1340	1300	1230
$i=6$	1450	1410	1370	1290
$i=7$	1530	1490	1440	1360
$i=8$	1600	1550	1510	1430
$i=9$	1680	1630	1580	1500

Табл. 5 – Нормативные потребности в энергии  $E_{ij}^M$  для мужчин, ккал/сут

	$j=1$	$j=2$	$j=3$	$j=4$
$i=1$	1450	1370	1280	1180
$i=2$	1520	1430	1350	1240
$i=3$	1590	1500	1410	1300
$i=4$	1670	1570	1480	1360
$i=5$	1750	1650	1550	1430
$i=6$	1830	1720	1620	1500
$i=7$	1920	1810	1700	1570
$i=8$	2010	1900	1780	1640
$i=9$	2110	1990	1870	1720

Согласно таблицам НФП [1, 2] применительно к витаминам и минеральным веществам суточная таблица указанных нутриентов не зависит от пола (за исключением железа), физической активности, а также от возраста (за исключением кальция). В таблице 6 приведены сточные потребности [мг] для женщин массой 60 кг и мужчин массой 70 кг.

Табл. 6 – Данные НФП в витаминах и минеральных веществах, мг

С	В <sub>2</sub>	РР	К	Р	Na	В <sub>1</sub>	А	каротин	Mg	Fe
90	1,8	20	1000	800	1300	1,5	900	5,0	400	18 (женщины) 10 (мужчины)

Суточная потребность в кальции (Ca) равна 1000 мг для возраста  $a \leq 59$  и 1200 мг для возраста  $a \geq 60$ .

Усвоение энергоресурсов и микроэлементов зависит от способа приготовления пищи, типа пищи (животная или растительная), сочетания продуктов. Соответствующие величины составляют содержание БДПБ на основе материалов [7, 8].

### 3.2. Среднегодовое потребление в целом

Потребляемые продукты питания подразделяются на 10 крупных категорий (см. таблицу 7). Иногда с целью детализации потребления хлеб, мясо птицы и крупы выделяют отдельно. Статистика суммарного потребления отвечает некоторому «среднему» человеку 38 лет (второй возрастной диапазон) массой 75 кг (шестой весовой диапазон) средней физической активности (коэффициент увеличения данных в таблицах 3 и 4 – в 1,8 раза). В таблице 7 приведены данные [9] среднегодового потребления продуктов по крупным категориям.

Табл. 7 – Среднегодовое потребление продуктов на человека, кг

Виды продуктов	2017 г.	2018 г.	Калорийность суточного рациона, %
Хлебные продукты (хлеб и макаронные изделия в пересчете на муку, мука, крупа и бобовые)	97,0	95,7	35,4
Картофель	59,4	58,7	3,6
Овощи и бахчевые	102,3	104,1	2,7
Фрукты и ягоды, включая сушеные в пересчете на свежие	73,0	73,7	3,7
Мясо животных и мясопродукты в пересчете на мясо	62,1	62,5	11,9
Мясо птицы	26,1	26,6	6,0
Рыба и рыбопродукты в пересчете на рыбу	21,5	21,7	1,7
Сахар, включая кондитерские изделия в пересчете на сахар	31,4	31,3	12,0
Масло растительное и другие жиры	10,8	10,7	9,4
Молоко и молочные продукты в пересчете на молоко	265,5	265,5	11,9
Яйца, штук	230	231	1,8

Переводя данные таблицы 7 в суточное потребление, получаем, что в среднем человек съедает за день примерно 265 г хлебопродуктов включая каши, 280 г овощей, 160 г картофеля, 200 г фруктов, 300 г мяса включая птицу и рыбу, 85 г сахара, 30 г растительного масла и 700 г молочной продукции. Согласно БДПБ, этому набору отвечает калорийность 2630 ккал, что для среднего уровня физической активности порядка 1,8 означает нормативное энергопотребление (таблицы 4 и 5) на уровне 1460 ккал. В целом эти данные приблизительно соответствуют параметрам «среднего» человека согласно таблицам 4 и 5, т.е. рекомендации Минздрава в среднем выполняются.

Также можно сделать вывод о том, что уровень потребления по категориям продуктов весьма стабилен, так что пищевые привычки населения предположительно постоянны.

Среднесуточный рацион по макронутриентам состоит из 80 г белков, примерно 50 г из которых животного происхождения, 108 г жиров (из них 69 г животного происхождения) и 335 г углеводов, в том числе 15 г животного происхождения. По калорийности этот средний рацион состоит из 320 ккал, отвечающих потреблению белков, 970 ккал жиров и 1340 ккал углеводов. Заметим, что потребление белков, жиров и углеводов должно находиться в пределах физиологически необходимых процентных соотношений по калорийности между ними: 12 %, 30 % и 58 % соответственно. Убеждаемся, что фактическое потребление составляет в процентном содержании несколько другие пропорции: 12 %, 37 %, 51 %.

Проведенный анализ, а также данные таблицы 7 будут использованы для разработки алгоритма создания типовых меню блюд (см. далее п. 5).

### *3.3. Среднегодовое потребление отдельных домохозяйств*

Если рассмотреть годовое потребление продуктов питания отдельными хозяйствами, то можно убедиться, что по категориям таблицы 7 потребление варьируется не сильно. Относительная погрешность среднего значения каждой из категорий продуктов составляет согласно [9] величину порядка 1 %.

Заметные отличия связаны с детализированным потреблением конкретных подкатегорий продуктов – определенного мяса или рыбы, круп, хлеба и т.д. При этом вариация калорийности суточного питания остается в пределах 1,5 %, тогда как относительное отклонение среднесуточного потребления конкретного хозяйства по определенным подкатегориям продуктов (мяса свинины или гречневой крупы и т.п.) может отличаться от среднего хозяйства на 100 %.

Существенно различаются между хозяйствами также и величины потребления микронутриентов. Вариации там могут достигать 100 % по сравнению со среднестатистическим хозяйством.

Это свидетельствует не только о региональных, культурных и социально-экономических различиях, но и о том, что однообразным меню, состоящим из одних и тех же продуктов по подкатегориям, нельзя обеспечить питание населения страны. Следовательно, необходимо обработать статистику потребления продуктов в отдельных домашних хозяйствах с тем, чтобы сформировать классы таких хозяйств по типовым продуктовым наборам и, как следствие, разработать соответствующие типовые меню. Создание типовых паттернов, представляющих собой частоты употребления тех или иных блюд, является основой для практического использования цифровой нутрициологии по разработке персонифицированных рекомендаций оптимального питания.

## **4. Математические модели в задачах нутрициологии**

### *4.1. Интерполяция параметров внутри класса*

В п. 2 были введены укрупненные классы биометрических параметров пользователя. Однако при разработке персонифицированных рекомендаций

такая сетка оказывается слишком грубой. Пользователи отличаются один от другого также и в пределах одного класса, в котором, например, уравниены люди возрастов от 18 до 29 лет с массами от 45 до 50 кг. Поэтому необходимо создать алгоритм, «размазывающий» рекомендации [1, 2] на более мелкий масштаб: по возрастам с шагом в 1 год, а по массе тела – с шагом в 1 кг. Из оценки точности статистики по формуле (2.1) следует, что создание такого алгоритма не может опираться на эмпирические данные, поскольку данных с нужной точностью просто не существует (по крайней мере в России): для этого потребовалось бы порядка 1 млрд респондентов. Тогда предлагается следующая модель интерполяции данных  $E_{ij}^{F,M}$ , приведенных в таблицах 4, 5.

Обозначим  $a_j$  значения возрастов в узлах исходной укрупненной сетки, и  $m_i$  значения масс (верхние индексы  $F, M$  для краткости опущены).

Рассмотрим для каждого  $j$  функцию  $e_j(m)$ , заданную в узлах  $m_i$  на отрезке  $[m_1, m_9]$  равенствами:

$$e_j(m_i) = E_{ij}. \quad (4.1)$$

Продолжим эту функцию по кусочно-линейной интерполяции в остальные точки отрезка  $[m_1, m_9]$  и вычислим ее среднее значение:

$$\langle e_j \rangle = \frac{1}{m_9 - m_1} \int_{m_1}^{m_9} e_j(m) dm. \quad (4.2)$$

Для каждого из четырех значений  $j$  можно найти такое значение аргумента  $m_j^*$  функции  $e_j(m)$ , для которого  $\langle e_j \rangle = e_j(m_j^*)$ . Это значение всегда существует в силу известной теоремы математического анализа о среднем значении определенного интеграла, поскольку  $e_j(m)$  – непрерывная монотонно возрастающая функция:

$$e_j(m) = \frac{(m_{i+1} - m)E_{ij} - (m_i - m)E_{i+1,j}}{m_{i+1} - m_i}. \quad (4.3)$$

Вычисляя эти значения  $m_j^*$  для таблиц 4 и 5 выясняем, что они не зависят от  $j$  и приближенно равны  $m^{*F} = 60$  кг для женщин и  $m^{*M} = 70$  кг для мужчин. Это есть критическое свойство матриц ВОО, позволяющее построить модель пересчета рекомендуемых величин  $E_k$  потребления  $k$ -го нутриента с узлов крупной сетки на мелкий масштаб по закону сохранения площадей.

Пусть заданы масса  $m$  и возраст  $a$  пользователя. Найдем номера  $1 \leq i \leq 8$ ,  $1 \leq j \leq 4$ , такие, что  $m \in [m_i, m_{i+1}]$ ,  $a \in [a_j, b_j]$ . Рассмотрим прямоугольник ABCD, вершины которого являются узлами крупной сетки, как показано на рис. 4.

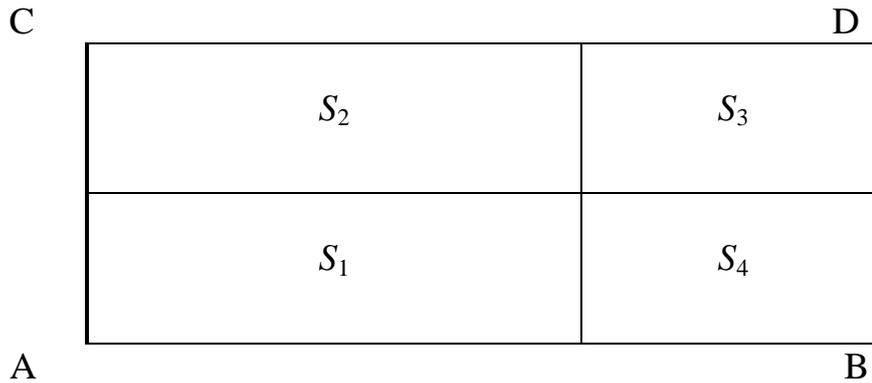


Рис. 4 – Преобразование ячейки параметров

Точка  $(a, m)$  лежит в прямоугольнике ABCD. Проходящие через эту точку вертикальная и горизонтальная прямые, параллельные соответствующим сторонам прямоугольника, делят его на четыре прямоугольника, площади которых обозначены  $S_1, S_2, S_3, S_4$ . Пусть  $S=S_1+S_2+S_3+S_4$ . Тогда норма  $E$  физиологической потребности в точке  $(a, m)$  равна:

$$E(y, m) = E(A) \cdot S_3/S + E(B) \cdot S_2/S + E(C) \cdot S_4/S + E(D) \cdot S_1/S, \quad (4.4)$$

$$S_1 = (a - a_j)(m_{i+1} - m), \quad S_2 = (a - a_j)(m - m_i),$$

$$S_3 = (b_j - a)(m - m_i), \quad S_4 = (b_j - a)(m_{i+1} - m),$$

а значения НФП в узлах сетки ( $E(A)$  и т.д.) известны. Учет группы физической активности дается умножением на соответствующий коэффициент. Тем самым значения НФП полностью вычисляются на сетке более мелкого масштаба.

#### 4.2. Индикаторы оптимальности питания

После того, как определен вектор  $E_i$  суточных потребностей пользователя в  $i$ -ом нутриенте (т.е. рекомендуемые значения НФП), требуется решить задачу алгоритмического подбора продуктовой корзины  $\{M_j\}$ ,  $j=1,2,\dots,n$ , состав которой может обеспечить такое потребление. Здесь  $M_j$  – масса  $j$ -го продукта (именно продукта, а не категории продуктов из таблицы 7). Сложность этой процедуры обусловлена следующими факторами.

Во-первых, искомое решение неоднозначно, ибо возможное количество  $K$  продуктов питания,  $n$  из которых отбираются в корзину, существенно превосходит число балансируемых нутриентов. Поэтому при одних и тех же значениях  $E_i$  могут быть различны как массы продуктов в собранной корзине, так и ее состав. Например, для каждой из 10 категорий продуктов существует как минимум несколько десятков подкатегорий (хлеб – 20 сортов, макаронные изделия – 20 сортов, крупы – 10 видов и т.д.). Произвольный выбор, допустим, трех категорий из возможных десяти реализуется примерно 100 способами ( $10!/(3!7!)=120$ ). Тогда для совокупности продуктов из 10 категорий получаем по крайней мере  $100^{10} = 10^{20}$  продуктовых корзин. Следовательно, простым перебором находить наилучшую из них не очень осмысленно. Заметим, что пока еще не указан и критерий наилучшей корзины.

Во-вторых, компьютерная программа должна позволять составить корзину на произвольный временной диапазон по желанию пользователя. Вследствие этого надо иметь алгоритм преобразования среднегодового потребления продуктов питания на другой промежуток времени. Это преобразование не может быть просто количественным пропорциональным изменением масс потребляемых продуктов, но должно затрагивать также и сам их перечень. Например, при планировании питания на неделю не получится включить в него несколько десятков видов рыб, мясо всех птиц и животных и т.д.

Поэтому требуется некоторая априорная модель подбора корзины в «нулевом приближении», которую можно за небольшое количество итераций привести к рекомендуемому персонализированному составу. Концепция такой модели изложена ниже в п. 5.

В-третьих, исходные данные по химическому составу продуктов имеют естественный статистический разброс, поэтому нет гарантии, что подобранная продуктовая корзина фактически будет обладать нужным составом нутриентов. Следовательно, необходимо разработать индикатор, который показывал бы, что достаточная и не улучшаемая далее точность в составе нутриентов достигнута, а также и то, что предлагаемый набор продуктов имеет достаточное разнообразие как по категориям (см. таблицу 7), так и по конкретным продуктам в подкатегориях.

Обозначим через  $A_{ij}$  нутриентную матрицу, элементы которой равны содержанию  $i$ -го нутриента в единице  $j$ -го продукта. Предположим, что распределения нутриентов в каждом продукте независимы и образуют совокупность нормально распределенных случайных величин со средним значением  $\mu_{ij}$  и дисперсией  $\sigma_{ij}^2$ . Тогда получаем, что нутриентный состав продуктовой корзины, определяемый формулой

$$X_i = \sum_{j=1}^n A_{ij} M_j, \quad (4.5)$$

есть нормально распределенная случайная величина со средним значением

$$\bar{X}_i = \sum_{j=1}^n \mu_{ij} M_j \text{ и дисперсией } \sigma_{X_i}^2 = \sum_{j=1}^n \sigma_{ij}^2 M_j^2.$$

Далее можно обсуждать, при каком наборе продуктов корзина устойчивее по  $i$ -му нутриенту. Индексом устойчивости называется нормированное среднее:

$$\xi_i = \frac{\bar{X}_i}{\sigma_{X_i}}. \quad (4.6)$$

Чем выше индекс устойчивости, тем меньше относительное отклонение фактической величины нутриента от своего среднего значения.

Однако корзина характеризуется многими нутриентами. Оптимизировать ее по каждому нутриенту может не получиться, потому что исходные

дисперсии  $\sigma_{ij}^2$  являются объективным свойством продукта и, как правило, не могут быть существенно изменены. Следовательно, надо разработать модель индикатора устойчивости корзины в целом. Кроме того, надо учесть, что балансируемые нутриенты эссенциальны, и потому отклонение каждого из них за пределы некоторых нормативных значений может привести (конечно, с некоторой вероятностью) к заболеваниям – разным для разных нутриентов. Какие из заболеваний считать более серьезными, а какие – менее, чтобы ранжировать превышения нормативов по степени опасности вероятных последствий, в настоящее время достоверно (то есть на уровне точной функциональной связи) не известно. Вследствие этого приходится экспертным образом вводить индивидуальные весовые коэффициенты  $w_i$  по отдельным нормированным индексам  $\xi_i$  так, чтобы взвешенные индексы  $w_i|\xi_i|$  были бы статистически и эссенциально равноценны. Построение таких коэффициентов – отдельная задача на стыке статистики, нутрициологии и медицины. В качестве ориентировочных значений можно предложить величины, полученные путем следующих рассуждений. Поскольку известны рекомендуемые пределы  $[\alpha_i, \beta_i]$  потребления каждого из нутриентов, а также известно и распределение величины (4.5), отвечающей построенной корзине продуктов, то можно вычислить вероятность того, что  $X_i \in [\alpha_i, \beta_i]$ . Обозначим эту вероятность  $p_i$ :

$$p_i = \frac{1}{\sqrt{2\pi}\sigma_{X_i}} \int_{\alpha_i}^{\beta_i} \exp\left(-\frac{(X_i - \bar{X}_i)^2}{2\sigma_{X_i}^2}\right) dx. \quad (4.7)$$

Если вероятность (4.7) близка к единице, то в повышении устойчивости потребления данного нутриента нет необходимости, и потому нормированная величина (4.6) незначительна при оптимизации корзины в целом. Если же эта вероятность мала, то вклад соответствующей нормированной величины должен быть большим. Основные два варианта введения весов тогда выглядят следующим образом. Либо мы считаем, что веса обратно пропорциональны найденным вероятностям (сумма весов равна единице), т.е.

$$w_i = \frac{1}{Wp_i}, \quad W = \sum_{i=1}^{15} 1/p_i, \quad (4.8)$$

либо они пропорциональны вероятности выхода величины  $X_i$  за пределы  $[\alpha_i, \beta_i]$ , т.е.

$$w_i = (1 - p_i)/14. \quad (4.9)$$

После того, как указанные весовые коэффициенты определены и выбрана модель взвешивания, можно строить модель оптимизации устойчивости продуктовой корзины. Естественным представляется максимизация суммы взвешенных индексов:

$$J = \sum_{i=1}^{15} w_i |\xi_i| \rightarrow \max, \quad (4.10)$$

или, как вариант,

$$J = \sqrt{\sum_{i=1}^{15} w_i \xi_i^2} \rightarrow \max. \quad (4.11)$$

Кроме введенного индикатора, существуют еще различные индексы [10] – индекс качества питания (ИКП), индекс здорового питания (ИЗП), индекс разнообразного питания (ИРП) и др., в которых также отражены различные данные о роли пищевых факторов в развитии алиментарно-зависимых хронических неинфекционных заболеваний (ХНИЗ). Это ранговые экспертные индикаторы, которые отражают точки зрения тех или иных диетологических систем. Поскольку в нашей работе ХНИЗ не обсуждаются, то индикаторы типа ИКП мы пока не включаем в процесс оптимизации.

Пусть теперь в результате некоторого подбора продуктовой корзины получены конкретные величины  $X_i$ , определяемые по формуле (4.5). Обозначим их отклонения от рекомендованных значений  $E_i$ , которым в идеале

должны быть равны средние значения  $\bar{X}_i = \sum_{j=1}^n \mu_{ij} M_j$ , через  $\delta_i$ :

$$\delta_i = \bar{X}_i - E_i = \sum_{j=1}^n \mu_{ij} M_j - E_i. \quad (4.12)$$

Тогда можно определить взвешенную относительную ошибку корзины по формуле

$$\Delta = \sum_{i=1}^{15} w_i |\delta_i| / E_i, \quad (4.13)$$

где  $w_i$  определена по формуле (4.9). Этот индикатор дополняет индекс устойчивости (4.10) или (4.11), поскольку предыдущие выражения не отражали точность выполнения рекомендаций по питанию. Здесь оптимизация корзины состоит в том, что выбирается тот вариант, при котором  $\Delta \rightarrow \min$ .

### 4.3. Демографическая модель

Поскольку целевой функцией нутрициологии является повышение здоровья людей и, как следствие, увеличение продолжительности жизни, то одной из важных моделей является формулировка зависимости ожидаемой продолжительности жизни от качества питания. Для построения такой модели надо использовать демографическое уравнение и дополнить зависимость коэффициентов, определяющих его решение, от индикаторов оптимального питания.

Горизонт планирования в данном случае есть продолжительность жизни, то есть величина, много большая, чем 1 год. Тогда вместо дискретной модели вида (2.2) можно использовать уравнение изменения численности возрастных групп в дифференциальной форме, а уравнение воспроизводства населения – в интегральной форме. Также имеет смысл основываться на базовом сценарии, предполагающем, что коэффициенты рождаемости и смертности по популяции

не меняются со временем. Этот подход призван оценить возможные изменения среднестатистической продолжительности жизни вследствие оптимизации структуры питания при прочих равных условиях.

Соответствующая система уравнений имеет вид [11]:

$$\frac{\partial N^{F,M}(a,t)}{\partial t} + \frac{\partial N^{F,M}(a,t)}{\partial a} = -q^{M,W}(a)N^{F,M}(a,t), \quad a > 0, t > 0;$$

$$N^{F,M}(0,t) = k^{F,M} \int_0^{\infty} B(a)N^F(a,t)da; \quad k^F = 0,488, \quad k^M = 0,512; \quad (4.14)$$

$$N^{F,M}(a,0) = N_0^{F,M}(a).$$

Эта система имеет точное аналитическое решение. Запишем его для краткости только для женской части населения, так как для мужской решение имеет аналогичный вид.

$$N^F(a,t) = re^{s(t-a)}Q^F(a), \quad Q^F(a) = \exp\left(-\int_0^a q^F(x)dx\right). \quad (4.15)$$

Величина  $Q^F(a)$  называется функцией дожития. Она представляет собой экспоненту от накопленной повозрастной функции смертности. Постоянные коэффициенты  $r$  и  $s$  в (4.15) определяются по начальному распределению и функции рождаемости. Вводятся функции:

$$R^F(a) = Q^F(a)B^F(a), \quad (4.16)$$

$$\varphi^F(a) = \int_0^{\infty} R^F(a+x) \frac{N_0^F(x)}{Q^F(x)} dx, \quad (4.17)$$

а также их образы после преобразования Лапласа  $\tilde{R}^F(p)$ ,  $\tilde{\varphi}^F(p)$ , где

$$\tilde{R}^F(p) = \int_0^{\infty} e^{-ap} R^F(a) da. \quad (4.18)$$

Величина  $s$  в (4.15) есть корень уравнения  $\tilde{R}^F(s) = 1$ ; в силу пологости функции  $Q(a)$  и унимодального вида функции рождаемости  $B(a)$  (рис. 2) это решение приближенно равно обратному возрасту матери в моде распределения (возрасту поколения)  $s = 1/a_m \approx 1/27$ . Постоянная  $r$  в (4.15) определяется начальным распределением согласно формуле

$$r = \lim_{p \rightarrow s} \frac{(p-s)\tilde{\varphi}^F(p)}{1 - \tilde{R}^F(p)}. \quad (4.19)$$

Из приведенного решения (4.15) следует, что ключевым демографическим параметром, определяющим продолжительность жизни, является накопленный коэффициент смертности. Именно на него может быть оказано влияние

посредством оптимизации системы питания. Важной моделью в этом смысле является модель стабильного населения, определяемая условием

$$\int_0^{\infty} B^{M,F}(a) Q^{M,F}(a) da = 1. \quad (4.20)$$

Для этой модели численность населения в каждой возрастной когорте постоянна, а само распределение по возрастам имеет вид

$$N(a) = N(0)Q(a). \quad (4.21)$$

Для коэффициентов смертности, представленных на рис. 3, модель стабильного населения имеет следующий вид (рис. 5):

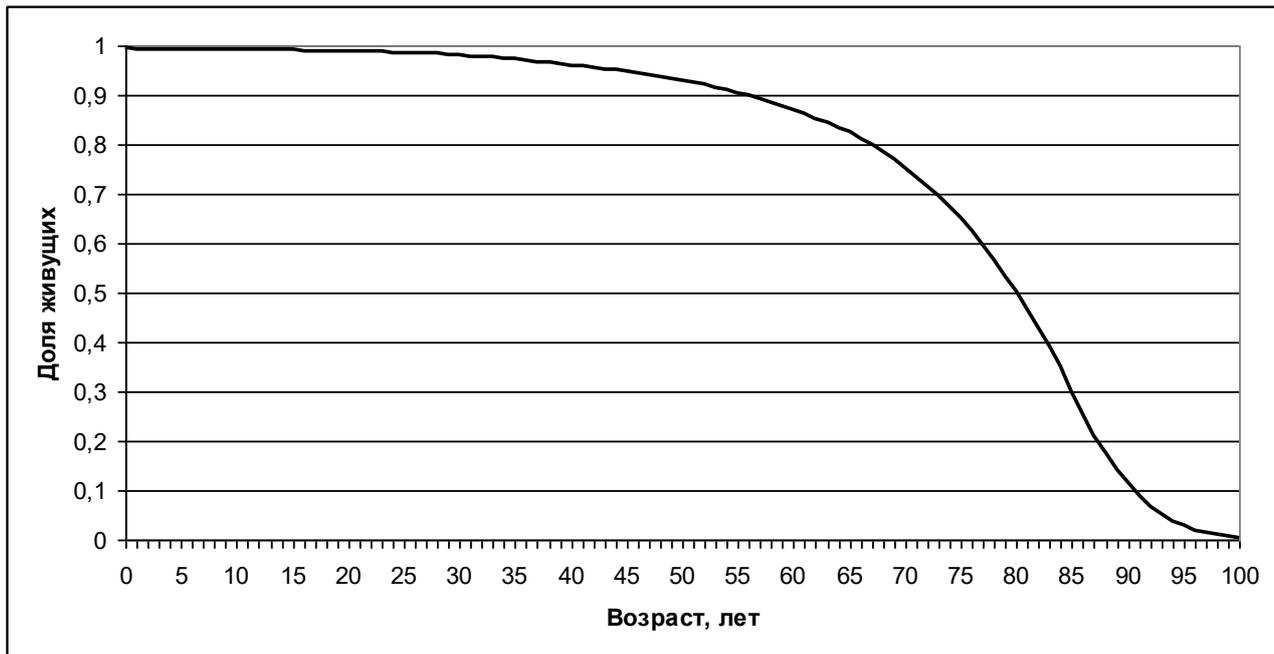


Рис. 5 – Распределение женщин по возрастам в модели стабильного населения для РФ

Сравнивая этот график с реальным распределением населения РФ (рис. 1), видим, что, если бы не внешние причины, в целом эта модель адекватно описывает основную тенденцию, связанную со старением. Тогда роль тех или иных действий, улучшающих здоровье населения, сводится в плане модели к снижению повозрастного коэффициента смертности, т.е. к поднятию кривой дожития, что означает увеличение вероятности дожить до определенного возраста. Опишем соответствующую модель в линейном приближении по отклонению от оптимальных рекомендаций применительно к продуктовой корзине.

Предположим, что при правильном питании, когда отклонение  $\delta_i$  в формуле (4.12) не выводит потребление данного нутриента за пределы  $[\alpha_i, \beta_i]$ , коэффициент смертности равен некоторой величине  $q(a)$ , а в случае превышения этих границ он возрастает в соответствии с тем, насколько по представлениям современной медицины сокращается в среднем жизнь от

развития сопутствующих заболеваний. Определим величину  $p'_i$  следующим образом. Если  $E_i + \delta_i > \beta_i$ , то

$$p'_i = \frac{1}{\sqrt{2\pi}\sigma_{X_i}} \int_{\alpha_i}^{E_i + \delta_i} \exp\left(-\frac{(X_i - E_i)^2}{2\sigma_{X_i}^2}\right) dx,$$

а если  $E_i - \delta_i < \alpha_i$ , то

$$p'_i = \frac{1}{\sqrt{2\pi}\sigma_{X_i}} \int_{E_i - \delta_i}^{\beta_i} \exp\left(-\frac{(X_i - E_i)^2}{2\sigma_{X_i}^2}\right) dx. \quad (4.22)$$

Мерой отклонения потребления  $i$ -го нутриента от оптимального будем считать величину  $\lambda_i = p'_i - p_i$ , где  $p_i$  определена в (4.7). Тогда

$$\lambda_i = \frac{1}{\sqrt{2\pi}\sigma_{X_i}} \int_{\beta_i}^{E_i + \delta_i} \exp\left(-\frac{(X_i - E_i)^2}{2\sigma_{X_i}^2}\right) dx,$$

либо

$$\lambda_i = \frac{1}{\sqrt{2\pi}\sigma_{X_i}} \int_{E_i - \delta_i}^{\alpha_i} \exp\left(-\frac{(X_i - E_i)^2}{2\sigma_{X_i}^2}\right) dx. \quad (4.23)$$

Пусть по оценкам средняя продолжительность жизни, определяемая как

$$\bar{a} = \int_0^{\infty} a d(1 - Q(a)) = \int_0^{\infty} a q(a) Q(a) da, \quad (4.24)$$

сократилась в силу неправильного питания по  $i$ -му нутриенту на  $a'_i$  лет в силу того, что коэффициент смертности возрос на величину  $c_i \lambda_i$  (для простоты равномерно во всех возрастах). Считая вклады от разных причин смертности суммирующимися, получаем связь средней продолжительности жизни с индикатором оптимального питания:

$$\bar{a} - \sum_{i=1}^{15} a'_i = \int_0^{\infty} a \cdot (q(a) + \Lambda) \cdot Q(a) e^{-\Lambda a} da, \quad \Lambda = \sum_{i=1}^{15} c_i \lambda_i. \quad (4.25)$$

Эта формула дает связь между эмпирическими данными в виде  $a'_i$  и  $\lambda_i$  и модельными коэффициентами  $c_i$ , если известны соответствующие увеличения коэффициентов смертности. Поскольку коэффициенты смертности малы, как и поправки к ним, то линейное приближение по  $\Lambda$  в (4.25) вполне адекватно. Тогда, разлагая экспоненту в ряд до члена первого порядка по  $\Lambda$ , получаем

$$a'_i = c_i \lambda_i \left( \sigma_a^2 + \bar{a}^2 - \kappa \right), \quad \kappa = \int_0^{\infty} a Q(a) da. \quad (4.26)$$

Коэффициент при  $c_i \lambda_i$  в (4.26) определяется по функции  $Q(a)$ . По данным за 2017 г. для населения России (в данном примере расчет для женщин)  $\bar{a} = 76$

лет, дисперсия продолжительности жизни равна  $\sigma_a^2 = 219$  (лет в квадрате), а интеграл  $\kappa = 3008$  (лет в квадрате), так что коэффициенты  $c_i$  оцениваются формулой

$$c_i = a'_i / (2987\lambda_i) \approx 3 \cdot 10^{-4} a'_i / \lambda_i. \quad (4.27)$$

#### 4.4. Модель баланса нутриентов

Поставим задачу определения величин порций продуктов (и блюд)  $M_j$  на основе известной нутриентной матрицы  $A_{ij}$  исходя из того, что нутриентный состав корзины продуктов (или меню блюд) в (4.5) равен рекомендуемым значениям НФП  $E_i$ :

$$\sum_{j=1}^K A_{ij} M_j = E_i. \quad (4.28)$$

При этом компоненты вектора решения (величины порций блюд) должны быть строго положительны.

Матрица  $A_{ij}$  представляет основную характеристику базы данных цифровой нутрициологии. Ее базисный минор определяет точность, с которой определяется весовой состав блюд меню. Формально решение уравнения (4.28) осуществляется средствами аппарата вычислительной линейной алгебры в зависимости от структуры матрицы  $A_{ij}$ . Однако на практике могут возникнуть вычислительные трудности, связанные с плохой обусловленностью.

Вариант 1. Число  $K$  блюд меню в точности равно числу балансируемых нутриентов. Матрица  $A_{ij}$  в этом случае – квадратная размеров  $15 \times 15$ . Если эта матрица невырождена, то существует обратная матрица  $A^{-1}$ , которая также считается известной, после чего решение уравнения (4.27) имеет вид

$$\mathbf{M} = A^{-1}\mathbf{E}. \quad (4.29)$$

Вариант 2. Число  $K$  блюд меню меньше числа балансируемых нутриентов. Тогда матрица  $A_{ij}$  – прямоугольная размеров  $15 \times K$ , базисный минор имеет ранг  $K < 15$ . Система в общем случае несовместна, т.е. получить решение задачи по всем нутриентам невозможно. Необходимое условие существования решения –  $\text{rang } A = \text{rang}[A, \mathbf{E}]$ , где  $[A, \mathbf{E}]$  – расширенная матрица, полученная из матрицы  $A$  добавлением вектора  $\mathbf{E}$  в качестве дополнительного столбца. В случае, когда это известное из курса линейной алгебры условие не выполнено, ищем обобщенное решение этой задачи, доставляющее минимум функции

$$\sum_{i=1}^{15} \left( \sum_{j=1}^K A_{ij} M_j - E_i \right)^2 \rightarrow \min, \quad 1 \leq j \leq K, M_j > 0. \quad (4.30)$$

Известно [12], что эта задача всегда имеет решение. Если при этом оказалось, что  $\text{rang } A = K$ , то решение единственное. Оно ищется из необходимого условия экстремума и имеет вид

$$\mathbf{M} = (A^T A)^{-1} A^T \mathbf{E}, \quad (4.31)$$

где  $A^T$  – транспонированная матрица, причем  $A^T A$  – положительно определенная и, в частности, невырожденная матрица.

**Вариант 3.** Число  $K$  блюд меню больше числа балансируемых нутриентов. Тогда базисный минор имеет ранг 15, а система (4.28) в общем случае имеет более одного решения. Представим матрицу  $A$  в блочном виде, выделив базисный минор  $G$  и оставшуюся подматрицу  $U$ :

$$A = \left\| \begin{array}{cc} G_{15 \times 15} & U_{15 \times K-15} \end{array} \right\|.$$

Тогда уравнение (4.28) имеет вид:

$$A\mathbf{M} \equiv G \begin{pmatrix} M_1 \\ \dots \\ M_{15} \end{pmatrix} + U \begin{pmatrix} M_{16} \\ \dots \\ M_K \end{pmatrix} = \begin{pmatrix} E_1 \\ \dots \\ E_{15} \end{pmatrix}. \quad (4.32)$$

Таким образом, коэффициенты использования первых 15 блюд меню зависят от произвольно выбранных неотрицательных коэффициентов остальных  $K - 15$  блюд при условии неотрицательности решения относительно всех компонент:

$$\begin{pmatrix} M_1 \\ \dots \\ M_{15} \end{pmatrix} = G^{-1} \begin{pmatrix} E_1 \\ \dots \\ E_{15} \end{pmatrix} - G^{-1} U \begin{pmatrix} M_{16} \\ \dots \\ M_K \end{pmatrix}. \quad (4.33)$$

Следовательно, алгоритм действий по составлению суточного меню состоит в выборе доминирующего по нутриентам базисного минора.

## 5. Базовые меню

Разработка базовых продуктовых наборов и соответствующих меню является главной составляющей компьютерной программы сопоставления режима питания конкретному пользователю. В основе лежит принцип иерархического построения БДПБ, где продукты сгруппированы в большие категории (см. таблицу 7), каждая из которых содержит подкатегории, для которых известны данные по химическому составу. Поскольку некоторые продукты являются одновременно блюдами (хлеб, макароны, крупы в виде каш, кондитерские изделия в пересчете на сахар), то с точки зрения алгоритмизации их удобно выделить в отдельные категории, которых, таким образом, становится 15, как и нутриентов: хлеб, макароны, крупы, остальные хлебобулочные изделия, мясо животных, мясо птиц, мясо рыб, картофель, овощи и бахчевые, фрукты и ягоды, сахар, остальные кондитерские изделия, масло растительное, молочные продукты, яйца.

Для выбора блюд меню из БДПБ удобно использовать двойное маркирование блюд, привязывающее их к категории блюда (суп, каша, мясное второе блюдо, салат, напиток и т.п.) и к формату приема пищи (завтрак, обед,

перекус, ужин). Выбор блюд естественно устроить по убыванию частоты их употребления в каждом формате приема пищи. Таким образом, чтобы реализовать модель выбора блюд, надо снабдить каждое блюдо, допустимое к приему пищи на завтрак, обед и т.д., частотой его суточного использования (т.е. эмпирической вероятностью). Для ее оценки используется как экспертный подход, состоящий в отборе определенных типов блюд, так и эмпирический, основанный на данных таблицы 7, из которых можно определить среднесуточное потребление продуктов по категориям. Однако при составлении меню нужны не категории продуктов, а категории блюд (некоторые совпадают с категориями продуктов).

Мы выделяем 15 основных категорий блюд: 1 – салат; 2 – горячая закуска; 3 – холодная закуска; 4 – первое блюдо; 5 – второе блюдо; 6 – гарнир; 7 – каша; 8 – приправы; 9 – молочные продукты и яйца; 10 – кисломолочные продукты; 11 – фрукты и сухофрукты; 12 – хлеб; 13 – остальные хлебобулочные изделия (кроме хлеба и макарон, включенных в гарнир); 14 – кондитерские изделия и сахар; 15 – напиток.

Далее экспертно выбирается формат меню, в котором устанавливаются категории блюд, употребляемых в данном приеме пищи. Например, обед, состоящий из двух блюд: это первое и напиток. Или обед, состоящий из восьми блюд: салат, первое, второе, гарнир, соус, хлеб, кондитерское изделие, напиток. Количество вариантов меню должно быть достаточным для создания разнообразного питания в зависимости от пожеланий пользователя.

С каждой категорией блюда, установленного в формате меню для каждого приема пищи, связывается подкатегория – собственно блюда. Например, категория «салат» включает подкатегории: 1.1 – из помидоров с луком репчатым с растительным маслом, 1.2 – из помидоров, огурцов и зелени со сметаной, 1.3 – оливье, 1.4 – из помидоров с салатом, курицей и яйцом, 1.5 – из капусты с яблоками и т.д.

Далее для каждого блюда устанавливается частота его использования на основе анализа годового потребления продуктов. Рассмотрим, например, получение оценки частоты употребления первых блюд. Из данных [9] следует, что мясо (животных, птиц и рыб) употребляется в пищу на уровне 72 кг в год или примерно по 200 г в сутки. Исходя из того, что в первом блюде собственно мяса содержится около 50 г, заключаем, что примерно 25 % мясной продукции идет на приготовление первых блюд. Из них 42 % – мясо животных, 37 % – мясо птиц, 21 % – мясо рыб. Кроме того, в качестве первого употребляются такие сезонные блюда, как окрошка на квасе, молочные супы, овощные супы-пюре. Долю таких блюд оценим в 10 % от общего количества первых блюд. В этом случае доля супов на мясном бульоне перенормируется на суммарную величину 90 %. Из них одну седьмую часть (т.е. примерно 14 %) отведем на рассольник, харчо и щи из квашеной капусты. Следовательно, если мы рассматриваем основные первые блюда, то в среднем получаются такие суточные частоты их употребления:

- окрошка  $p = 0,04$ ;

- холодный молочный суп  $p = 0,02$ ;
- овощной суп-пюре  $p = 0,04$ ;
- рыбный суп  $p = 0,16$ ;
- рыбная солянка  $p = 0,03$ ;
- «куриный» суп  $p = 0,28$ ;
- щи с «курицей»  $p = 0,05$ ;
- харчо  $p = 0,02$ ;
- мясная солянка  $p = 0,03$ ;
- суп на мясном бульоне  $p = 0,11$ ;
- щи с мясом  $p = 0,08$ ;
- щи с мясом и с квашеной капустой  $p = 0,03$ ;
- борщ  $p = 0,11$ .

Наименьшей частотой употребления в этом перечне является величина  $p_{\min} = 0,02$ . Наибольшая частота употребления равна  $p_{\max} = 0,28$ . Следовательно, минимальное количество дней, для которых можно составить меню, включающее все перечисленные блюда, равно  $T = 1/p_{\min} = 50$ . В течение этого времени пользователю будет предложено меню, в котором каждое первое блюдо используется в количестве  $pT$  раз (естественно, округление проводится до ближайшего целого числа). Так, куриный суп будет предложен 14 раз, борщ 6 раз, суп на мясном бульоне 6 раз, щи 6 раз (в том числе из квашеной капусты 2 раза) и т.д. Если требуется составить план на меньшее число дней  $T' < T$ , то выбираются частоты использования блюд, превосходящие  $p'_{\min} = 1/T'$ , начиная от наибольшей частоты  $p_{\max}$ . Если  $T' \leq 1/p_{\max} \approx 3$ , то все суточные меню содержат одно и то же первое блюдо. Если  $T' > 1/p_{\max}$ , то начинают добавляться блюда в соответствии с убыванием их оцененных частот.

Если  $T' > T$ , то первые  $n = [T'/T]$  временных промежутков по 50 дней повторяют составленное на 50 дней меню, а остаток, который меньше  $T$  дней, планируется по вышеописанному способу, когда  $T' < T$ .

Аналогично рассматриваются, например, гарниры. Исходя из величины порции гарнира в 150 г для круп, 200 г для картофеля, овощей и макаронных изделий, в соответствии с данными [9] получаем, что рис и макароны употребляются как блюдо 1 раз в неделю (т.е. суточная частота их потребления равна примерно 0,14), все остальные крупы и бобовые – 2 раза в неделю, картофель 1 раз в день, овощи 2 раза в день.

Хлеб употребляется ежедневно в количестве 4 порционных кусков пшеничного и 2 кусков ржаного.

Таким же способом определяются априорные частоты употребления и других категорий блюд. В пределах одного дня каждое блюдо выбирается только один раз.

В результате БДПБ структурируется на подразделы, каждый из которых отвечает определенному приему пищи в течение суток и в которых содержатся варианты типовых меню.

В работе [13] была описана клиент-серверная организация СУБД для реализации решения данной задачи цифровой нутрициологии по выбору меню блюд и расчета величины их порций. В качестве системы хранения данных была выбрана объектно-реляционная СУБД [14] с открытым исходным кодом PostgreSQL. Отметим, что PostgreSQL обладает рядом преимуществ по сравнению с обычными реляционными базами данных. Одним из них является поддержка широкого спектра структур и типов данных: uuid, денежного, перечисляемого, геометрического, бинарного типов, сетевых адресов, битовых строк, текстового поиска, xml, json, массивов, композитных типов и диапазонов, а также некоторых внутренних типов для идентификации объектов.

Кроме большого количества поддерживаемых структур в PostgreSQL предусмотрена интеграция с языками программирования C++, Python, Perl.

Поддержка такого количества структур упрощает работу с базой данных и улучшает гибкость использования такой СУБД в задачах машинного обучения и при разработке программного обеспечения.

После того, как построена программная среда для проведения вычислений, можно перейти к численной реализации формул (4.29), (4.33).

## 6. Расчет величин порций блюд

Предположим, что на основе построенного перечня блюд конкретного меню и заданного вектора НФП делается расчет величин порций на каждый день. Для простоты рассмотрим здесь ситуацию, когда число блюд точно равно числу балансируемых нутриентов.

Меню на день: 1 – творог со сметаной и медом, 2 – хлеб пшеничный с сыром и сливочным маслом, 3 – чай, 4 – салат из свежих огурцов и помидоров с зеленым луком и оливковым маслом, 5 – суп-щи зеленые со сметаной и яйцом, 6 – хлеб ржаной, 7 – картофель жареный, 8 – мясная поджарка из говядины, 9 – шоколадная конфета, 10 – компот из сухофруктов, 11 – апельсин, 12 – кофе, 13 – печенье, 14 – каша гречневая на молоке, 15 – булочка с повидлом.

Матрица  $A_{ij}$  в (4.28) невырождена, но, проведя формальные вычисления по формуле (4.29), получаем в качестве решения вектор, часть компонент которого отрицательна, что не имеет смысла, а некоторые величины порций аномально велики.

Первая причина такого поведения решения состоит в том, что некоторые нутриенты малых концентраций присутствуют только в одном каком-нибудь блюде, так что его приходится взять настолько много, что другое блюдо для баланса уже других нутриентов приходится брать с отрицательным весом. Следовательно, суточный расчет несостоятелен из-за невозможности каждый

день реализовывать в питании средние нормативы, полученные усреднением за год. Поэтому меню надо составлять на достаточно большой промежуток времени, как указывалось в п. 5. После этого вместо исходной задачи решается уравнение относительно величин порций для меню на горизонт планирования  $T \geq 50$  неделю, а затем проводится сопоставление этих блюд по дням с учетом уже имеющихся суточных меню.

Однако этот подход еще не полностью устраняет вычислительные артефакты. Вторая причина состоит в том, что нутриентная матрица  $A$  для выбранного меню является хоть и невырожденной, но плохо обусловленной, а элементы базы данных по химическому составу блюд известны с точностью порядка 10-15 %, что существенно больше, чем требуемая точность элементов матрицы для разделения точек спектра. Спектральный портрет матрицы  $A$  указанного суточного меню приведен на рис. 6.

Построение спектральных портретов матриц основано на процедуре, описанной в книге [15]. Портрет показывает области, в которых лежат собственные значения матриц, если элементы этих матриц известны с точностью, десятичный показатель степени которой указан в легенде.

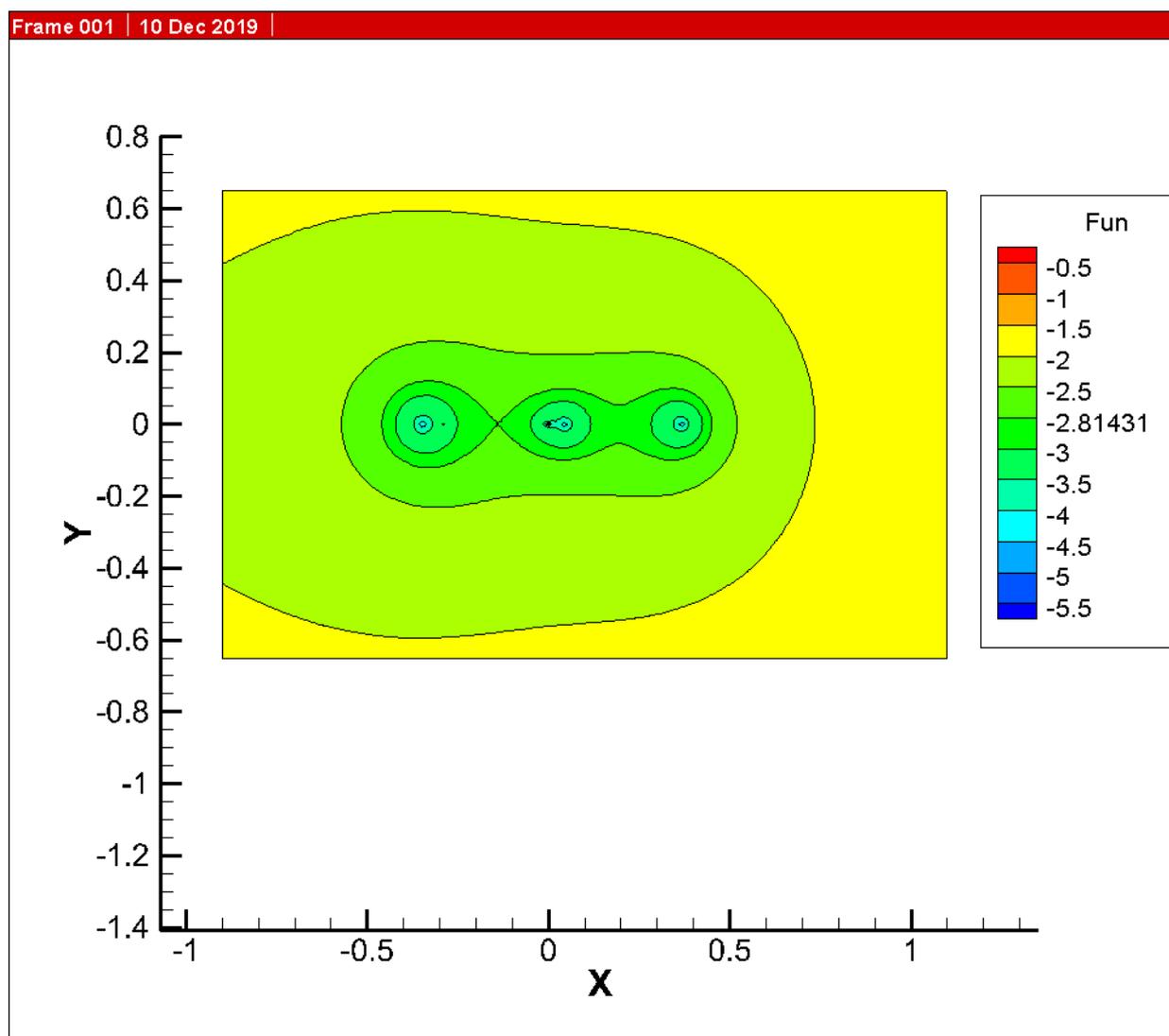


Рис. 6 – Спектральный портрет суточной нутриентной матрицы

Из рис. 6 следует, что для корректного решения уравнения (4.28) элементы матрицы  $A$  должны быть известны по крайней мере с точностью  $10^{-4}$ , что на самом деле не реализуется. Из этого рассмотрения следует, что увеличение количества дней при расчете меню не изменит долевого соотношения между нутриентами, и потому одновременный расчет величины порций при условиях контроля всех 15 нутриентов невозможен.

Поступим тогда следующим образом. Выберем строки системы (4.28) с доминирующими элементами, относящимися к макронутриентам, и проведем объединение блюд меню так, что вместо плохо обусловленной матрицы  $15 \times 15$  получилась бы «хорошая» матрица  $3 \times 3$  (рис. 7).

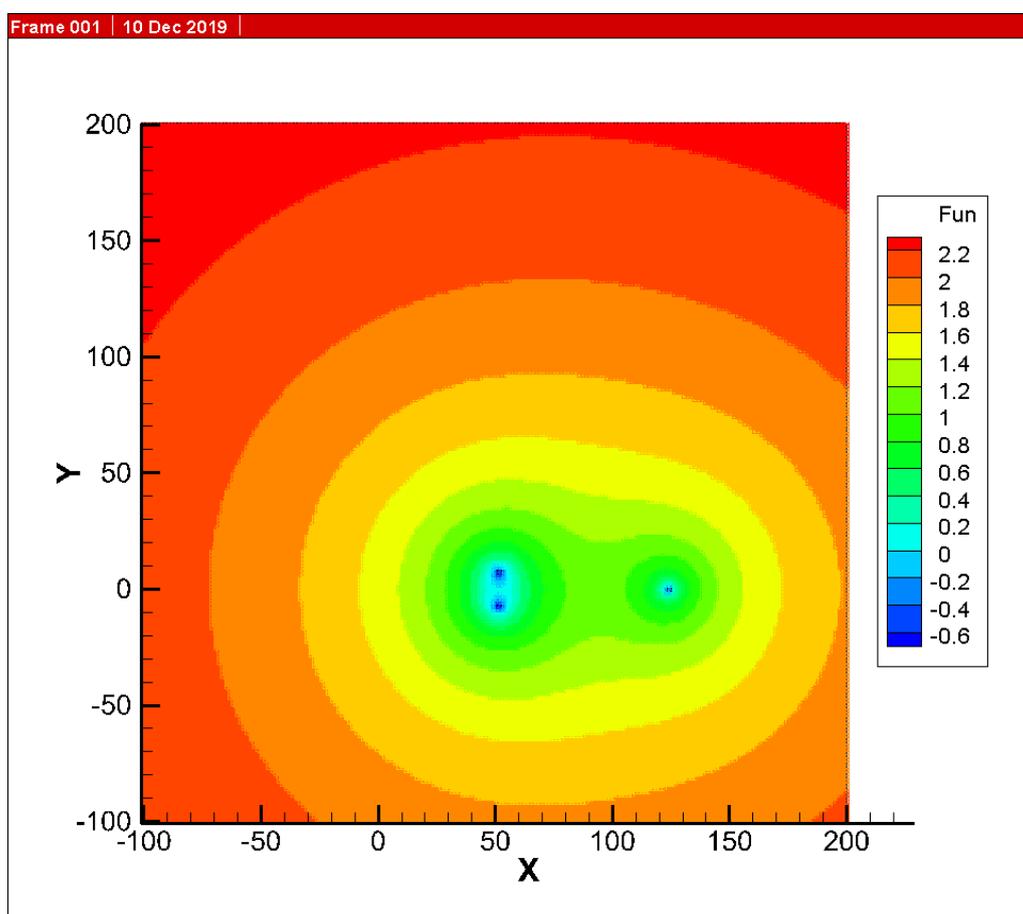


Рис. 7 – Спектральный портрет нутриентной матрицы для макронутриентов

Из рис. 7 видно, что собственные значения идентифицируются начиная с точности примерно 0,4 ( $10^{-0,4} \approx 0,4$ ). Это означает, что даже для весьма неточных измерений химического состава блюд решение задачи для макронутриентов будет корректным. Результаты расчета калорийности суточного меню приведены на рис. 8.

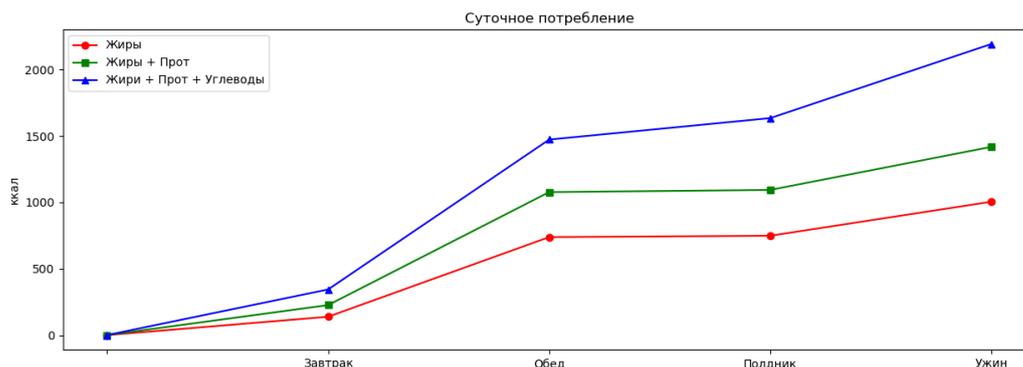


Рис. 8 – Суточное потребление основных нутриентов: жиров, белков и углеводов

Оставшиеся 12 нутриентов определяются после исключения из системы (4.27) блюд, дающих основной вклад в калорийность питания, величины порций которых были определены из решения системы  $3 \times 3$ . Соответствующее потребление всех нутриентов в сравнении с нормой приведено на рис. 9.

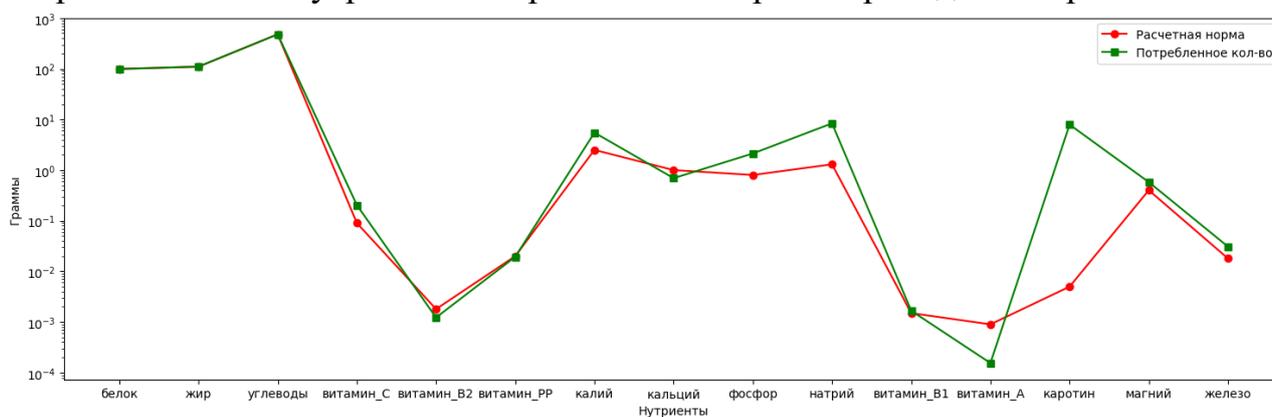


Рис. 9 – Суточное потребление нутриентов по расчету макронутриентов

Таким образом, предложенный двухступенчатый метод решения задачи определения величин порций меню позволил решить эту задачу.

## 7. Заключение

Данная работа имела целью сформулировать основные подходы к решению задач в области цифровой нутрициологии, разработать необходимый инструментарий и методологию, а также описать структуру эмпирических данных, которые необходимо использовать для создания численных алгоритмов применительно как к оптимизации питания больших групп людей по усредненным показателям, так и выдаче персонализированных рекомендаций по биометрии конкретного пользователя.

Большое внимание в работе уделено вопросам систематизации параметров, которыми описываются пользователи, и структурированию базы данных по химическому составу продуктов питания. Поскольку работа направлена на

создание концепции программного продукта, была проанализирована точность постановок задач и их решений.

Впервые на математическом уровне строгости сформулированы подходы к решению следующих задач: кластеризации данных в области питания, интерполяции рекомендованных значений ВОО на персональные данные пользователя, связи оптимального питания и средней продолжительности жизни, оценки устойчивости продуктовой корзины, а также формализации алгоритмического составления меню продуктов и блюд.

Дальнейшие направления исследований по анализу данных связаны с анализом чувствительности решений по входным параметрам, что необходимо для более полного представления о точности и корректности предлагаемых моделей.

Кроме того, необходимо будет разработать модели оптимизации продуктовой корзины в зависимости от медицинских, региональных и социально-экономических аспектов и построить соответствующие критерии оптимальности, включающие в том числе индексы здорового питания и им аналогичные.

Также важным направлением цифровизации является создание специальных паттернов – параметрических портретов пользователей типовых классов, для которых можно будет формировать базовые меню.

Описанные задачи представляют как теоретический интерес, так и практическую важность с точки зрения решения проблем продовольственной безопасности.

### **Список источников**

1. Методические рекомендации МР 2.3.1.2432-08 «Нормы физиологических потребностей в энергии и пищевых веществах для различных групп населения РФ» (утв. главным государственным санитарным врачом РФ 18 дек. 2008 г.).
2. Письмо Министерства здравоохранения РФ от 5 декабря 2014 г. N 13-2/1664 по МКБ десятого пересмотра.
3. Орлов Ю.Н. Оптимальное разбиение гистограммы для оценивания выборочной плотности распределения нестационарного временного ряда // Препринты ИПМ им. М.В. Келдыша. 2013. № 14. 26 с.
4. Демографический ежегодник России. 2017: Стат. сб./ Росстат. – М.: 2017. – 263 с.
5. Global Database on Body Mass Index / World Health Organization / [www.who.int](http://www.who.int)
6. Колосницына М.Г., Бердникова А.Н. Избыточный вес: сколько это стоит и что с этим делать? // Прикладная эконометрика, 2009. № 3. С. 72-93.
7. Тутельян В.А. и др. Картоoteca блюд диетического (лечебного и профилактического) питания оптимизированного состава. – М.: Национальная ассоциация клинического питания, 2008. – 448 с.
8. Тутельян В.А. Химический состав и калорийность российских продуктов питания. Справочник. – М.: ДеЛи принт, 2012. – 284 с.

9. Потребление продуктов питания в домашних хозяйствах // [www.old.gks.ru](http://www.old.gks.ru)
10. Мартинчик А.Н. Индексы качества питания как инструмент интегральной оценки рациона питания // Вопросы питания, 2019. Т. 88. № 3. С. 5-12.
11. Орлов Ю.Н., Суслин В.М. Кинетические уравнения для некоторых моделей демографии // Математическое моделирование, 2003. Т.15. №3. С.43-54.
12. Алексеев В.М. и др. Оптимальное управление. – М.: Наука, 1979. – 429 с.
13. Гавриков М.Б., Кислицын А.А., Орлов Ю.Н., Пестрякова Н.В. Вычислительные аспекты задач цифровой нутрициологии // Препринты ИПМ им. М.В. Келдыша. 2019. № 136. 16 с.  
URL: <http://library.keldysh.ru/preprint.asp?id=2019-136>
14. PostgreSQL <https://www.postgresql.org/>
15. Годунов С.К. Современные аспекты линейной алгебры. – Новосибирск: Научная книга, 1997. – 388 с.