

Н. Н. Яхно^{1, 2}, д-р мед. наук, проф., академик РАН, e-mail: info@ditc.ras.ru,

В. Н. Гридин¹, д-р техн. наук, проф., e-mail: info@ditc.ras.ru,

Д. С. Смирнов¹, канд. экон. наук, ст. науч. сотр., e-mail: info@ditc.ras.ru,

В. С. Панищев¹, канд. техн. наук, ст. науч. сотр., e-mail: info@ditc.ras.ru,

В. А. Парфенов², д-р мед. наук, проф., e-mail: info@ditc.ras.ru,

Т. М. Остроумова^{1, 2}, канд. мед. наук, мл. науч. сотр., e-mail: info@ditc.ras.ru,

Н. Н. Коберская^{1, 2}, канд. мед. наук, ст. науч. сотр., e-mail: info@ditc.ras.ru,

¹ ФГБУН Центр информационных технологий в проектировании РАН,

г. Одинцово Московской области,

² ФГАОУ ВО Первый Московский государственный медицинский университет

имени И. М. Сеченова Министерства здравоохранения Российской Федерации

(Сеченовский университет), г. Москва

Статистическая обработка и методика сокращения размерности пространства данных пациентов при анализе когнитивных нарушений¹

Предложено развитие методов статистической обработки информации и алгоритмов преобразования и анализа медицинских данных пациентов в целях обнаружения факторов субъективного и легкого когнитивного снижения. Показывается неприменимость классических классификационных моделей для таких данных с низкой исходной согласованностью. Предлагается методика выявления групп показателей, обладающих наибольшей согласованностью и разделительной способностью для априорных групп пациентов. Преобразованные в соответствии с методикой данные могут являться исходными данными для построения эффективных предиктивных моделей регрессионного типа.

Ключевые слова: когнитивные нарушения, сокращение размерности пространства, многомерные статистические методы, логистическая регрессия, метод главных компонент

Введение

В настоящее время информационные технологии в медицине активно развиваются и позволяют улучшить эффективность работы здравоохранительных учреждений. Медицинские информационные системы должны обеспечивать обработку больших объемов разнородных данных, что требует развития и использования новых современных методов.

Одной из проблем, связанных с обработкой данных в медицине, полученных от разных источников, является их несогласованность. Наиболее известные алгоритмы оценки и устранения несогласованности — алгоритмы DerSimonian-Laird [1] и Mandel-Paule [2] и их модификации [3—5] — предполагают нормальность

распределений погрешностей и проведение параметрических тестов, связанных с этим типом распределения. Это делает не всегда возможным их использование и, в частности, не позволяет применять эти методы для обработки данных медицинских исследований, не подчиняющихся известным теоретическим распределениям.

В настоящем исследовании понятие согласованности данных сформулировано на основе работ ИГМ СО РАН [6]. Пусть существует множество объектов $\mathcal{S} = \{S_1, \dots, S_m\}$ и набор функций $X = \{X_1, \dots, X_n\}$, определенных на множестве \mathcal{S} и принимающих количественные значения. Также имеется набор источников данных $\theta = \{\theta_1, \dots, \theta_p\}$ о значениях функции. Значения функции X_j , полученные на объектах S_1, \dots, S_m от источника θ_p , обозначены как $X_j(S_1, \dots, S_m)_p$. Данными называют совокупность подмножества объектов $\hat{\mathcal{S}} \subset \mathcal{S}$, подмножества функций $\hat{X} \subset X$ и их полученные значения от набора источников $\hat{\theta} \subset \theta$, которые могут быть объединены в кортежи вида

¹ Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проект № 19-29-01112 МК.

$$\langle X_j | \hat{S} = \{S_{i(1)}, \dots, S_{i(k)}\} \| X_j(S_{i(1)}, \dots, S_{i(k)}) \|_p \rangle.$$

Данные считают слабо согласованными, если в них присутствуют такие кортежи, что для двух источников $\theta_p, \theta_q \in \hat{\Theta}$, представления ими значений функции X_j на объектах \hat{S} отличаются не менее чем на определенную величину δ_j :

$$\| \|X_j(S_{i(1)}, \dots, S_{i(k)}) \|_p - \|X_j(S_{i(1)}, \dots, S_{i(k)}) \|_q \| > \delta_j.$$

Данные являются плохо согласованными, если в них присутствуют такие кортежи, что для двух источников $\theta_p, \theta_q \in \hat{\Theta}$, представления ими значений функции X_j на объектах \hat{S} отличаются не менее чем на определенную величину $\varepsilon_j > \delta_j$:

$$\| \|X_j(S_{i(1)}, \dots, S_{i(k)}) \|_p - \|X_j(S_{i(1)}, \dots, S_{i(k)}) \|_q \| > \varepsilon_j.$$

Величины δ_j, ε_j выбираются из содержательных соображений и зависят от проблемной области.

Одной из сложных и нерешенных задач является статистический анализ разнородных данных, получаемых при анализе когнитивного состояния пациента. Для решения данной задачи в настоящей работе предлагается доработать известный метод [6], который в первоначальном виде не применим для анализа данных о когнитивном состоянии вследствие невозможности однозначного определения набора функций и источников. Рассмотрим разработанный метод обработки данных пациентов в целях обнаружения факторов, влияющих на когнитивные нарушения.

Когнитивные нарушения по степени тяжести подразделяют на тяжелые (деменция) и додементные когнитивные расстройства. Внутри последних выделяют умеренные когнитивные нарушения (КН), а также легкое и субъективное когнитивное снижение [7]. Пациенты с субъективным когнитивным снижением (СКС) имеют жалобы когнитивного характера (например, на забывчивость, ухудшение концентрации внимания и памяти и др.) в отсутствие объективных признаков когнитивных нарушений. Под легким когнитивным снижением (ЛКС) подразумевают наличие у больного жалоб, в то время как результаты расширенного нейропсихологического тестирования не выявляют отклонений от возрастной нормы, достигающих степени умеренных когнитивных расстройств [7]. СКС и ЛКС представляют собой наиболее раннее проявление когнитивного снижения и повышают как риск, так и скорость развития деменции в дальнейшем [8].

Неотъемлемой и важной задачей является развитие методов эффективного преобразования статистических данных о состоянии па-

циентов для ответа на две группы вопросов: какие физиологические характеристики наибольшим образом влияют на риск развития когнитивных нарушений и какие характеристики следует использовать для объективного разделения пациентов с жалобами когнитивного характера на подгруппы с СКС и ЛКС. Для решения данной задачи разработана методика применения методов статистической обработки данных с использованием современных информационных средств и технологий при выделении наиболее информационных критериев для исходных данных. Применяемые в работе методы комбинаторного анализа для бинарных априорных групп позволили добиться значительного уменьшения влияния несогласованности значений нейропсихологических тестов в разных группах, что характерно для первичных разнородных данных.

1. Постановка задачи.

Описание исходных данных

Рассмотрим следующую математическую постановку задачи. Пусть $S = \{S_1, \dots, S_m\}$, — выборка пациентов, а $X = \{X_1, \dots, X_n\}$ — множество функций, характеризующих процессы когнитивного нарушения. Сложность рассматриваемой проблемной области заключается в том, что набор таких функций и их определение на множестве S заранее не известно. Однако известен набор нейропсихологических тестов $\Theta = \{\theta_1, \dots, \theta_p\}$, которые исполняют роль источников данных о значениях набора функций $\tilde{X} \subset X$ для пациентов, состав которых частично и, нередко лишь на качественном уровне, известен. Обозначим $\|S_1, \dots, S_m | \tilde{X}_p \|_p$ — набор результатов теста θ_p для выборки пациентов S_1, \dots, S_m , являющихся оценкой некоего частично известного набора процессов когнитивного снижения. Таким образом, мы можем говорить о данных как о наборе кортежей следующего вида:

$$\langle \theta_p | S = \{S_1, \dots, S_m\} \| S_1, \dots, S_m | \tilde{X}_p \|_p \rangle.$$

Тогда слабо согласованные данные можно определить как данные, где для двух тестов θ_p и θ_q выполняется следующее неравенство:

$$\sum_{i=1}^m |\Psi_{pq}(\|S_i | \tilde{X}_p \|_p) - \Psi_{qp}(\|S_i | \tilde{X}_q \|_q)| > \delta_j,$$

где Ψ_{pq} и Ψ_{qp} — нормировочные функции перевода для сопоставления тестов. Стоит отметить, что выбор нормировочных функций также является отдельной задачей и может повлиять на точность решения. В качестве простейшего варианта использована ранговая

характеристика значения теста и осуществлен переход к оценке несогласованности через ранговые коэффициенты корреляции, или ранговые коэффициенты конкордации Кенделла, для всего набора тестов.

Введем оценку несогласованности относительно априорных групп. Введем функцию A_p , принимающую значения $\{1,2,3\}$, которая показывает на основании теста θ_p , к какой из трех групп относится пациент — группе контроля, группе с субъективными когнитивными нарушениями или группе с легким когнитивными нарушениями. Тогда меру несогласованности можно рассчитать как

$$D = 2C_p^2 - \sum_{p=1}^P \sum_{q=1}^P \left(\frac{\#\{S|A_p(\|S|\tilde{X}_p\|_p) = 1 \& A_q(\|S|\tilde{X}_q\|_q) = 1\}}{\#\{S_1, \dots, S_m\}} + \frac{\#\{S|A_p(\|S|\tilde{X}_p\|_p) = 2 \& A_q(\|S|\tilde{X}_q\|_q) = 2\}}{\#\{S_1, \dots, S_m\}} + \frac{\#\{S|A_p(\|S|\tilde{X}_p\|_p) = 3 \& A_q(\|S|\tilde{X}_q\|_q) = 3\}}{\#\{S_1, \dots, S_m\}} \right), p \neq q,$$

где # — мощность соответствующего множества.

С учетом того, что тесты θ являются интегральными характеристиками, каждый из которых оценивает множество процессов, влияющих на когнитивное смещение, несогласованность данных говорит об отсутствии набора доминантных процессов, общих для большинства пациентов, или о невозможности такие процессы выявить. В то же время, существующие исследования позволяют говорить о наличии влияния физиологических параметров на когнитивное состояние. Например, в работах [9–11] показано, что одним из самых главных факторов риска развития когнитивного снижения является артериальная гипертензия (АГ). Соответственно, необходимо перейти к такому формату данных, где несогласованность данных не будет помехой для нахождения статистически значимой взаимосвязи между доминантными когнитивными и физиологическими процессами.

Стоит отметить, что в контексте решаемой задачи и трактовки наличия несогласованности в данных разница между слабо согласованными данными и плохо согласованными данными несущественна.

В исследование были последовательно включены 102 пациента с АГ (средний возраст $51,2 \pm 5,7$ лет, 56 женщин, 46 мужчин) в возрасте от 40 до 59 лет на момент включения в исследование и 50 практически здоровых людей (средний возраст $47,3 \pm 5,2$ лет, 33 женщины,

17 мужчин) с нормальным уровнем артериального давления (АД) (контрольная группа). Пациенты с АГ были дополнительно разделены на две подгруппы: пациенты с СКС (средний возраст $52,3 \pm 5,1$ лет, 27 женщин и 10 мужчин) и ЛКС (средний возраст $50,6 \pm 5,9$ лет, 29 женщин и 36 мужчин).

Когнитивные функции оценивались с помощью теста "10 слов" [12], Монреальской шкалы оценки когнитивных функций (MoCA) [13, 14], частей А и В теста построения маршрута (Trail making test, TMT) [15], теста Струпа [16], теста вербальных ассоциаций (литеральные и категориальные ассоциации) [17].

Для проведения статистических расчетов использовали программное обеспечение R версии 3.6.1 (2019-07-05) и графическую среду RStudio.

2. Методы решения.

Статистический анализ исходных данных

На первом этапе рассматривались различия между контрольной группой и группой с пациентами с АГ. В табл. 1 представлены нейро-

Таблица 1

Значимо различающиеся признаки и степени их различия и вариации

Нейropsychологический тест	Относительная разница средних	Относительная разница медиан	Коэффициент вариации, контрольная группа	Коэффициент вариации, пациенты с АГ
TMT.B -TMT.A	41,80 %	47,75 %	34,92 %	50,85 %
Тест Струпа, коэффициент интерференции T3—T2*	27,25 %	15,22 %	48,36 %	67,96 %
TMT.B	23,01 %	19,17 %	25,55 %	36,94 %
Литеральные ассоциации	-18,22 %	-18,52 %	20,10 %	31,14 %
Тест Струпа, T3	17,01 %	5,71 %	27,72 %	37,52 %
Тест Струпа, T1	11,45 %	9,47 %	14,67 %	15,25 %
Категориальные ассоциации	-11,42 %	-6,67 %	23,26 %	37,32 %
Тест "10 слов". Отсроченное воспроизведение	9,94 %	27,27 %	30,57 %	25,08 %
Тест Струпа, T2	8,71 %	10,92 %	18,66 %	19,03 %
TMT.A	-4,36 %	-6,33 %	32,17 %	34,74 %
Тест MoCA	-2,37 %	0,00 %	5,71 %	5,23 %
Тест "10 слов". Непосредственное воспроизведение	-2,06 %	0,00 %	16,89 %	24,90 %
T3—T2 — разность затраченного времени на выполнение частей 2 и 3 теста Струпа				

психологические тесты, обладающие значимыми различиями, отсортированные в порядке убывания относительной разницы средних по группе контроля и группе пациентов с АГ. Для оценки значимости использовался тест Манна — Уитни, так как большая часть параметров не распределена по нормальному закону. Также в табл. 1 представлена относительная разница между медианами и коэффициенты вариации каждой из групп пациентов.

При анализе результатов нейропсихологического тестирования статистически значимо различались (прошли тест Манна—Уитни с достоверностью $p\text{-value} < 0,05$) только разница между временем выполнения частей В и А ТМТ и средний балл по тесту МоСА. Корреляционный анализ (рассчитывались как линейные коэффициенты корреляции Пирсона, так и ранговые коэффициенты Спирмена и Кенделла) показал отсутствие значимой статистической связи между любыми парами признаков. Также был рассчитан коэффициент конкордации Кенделла для всех шкал, использовавшихся для оценки когнитивных функций, дающий характеристику согласованности. Полученное значение составляет $W = 0,1233$ при $p\text{-value} = 0,0001159$, что говорит о низкой согласованности в структуре исходного пространства.

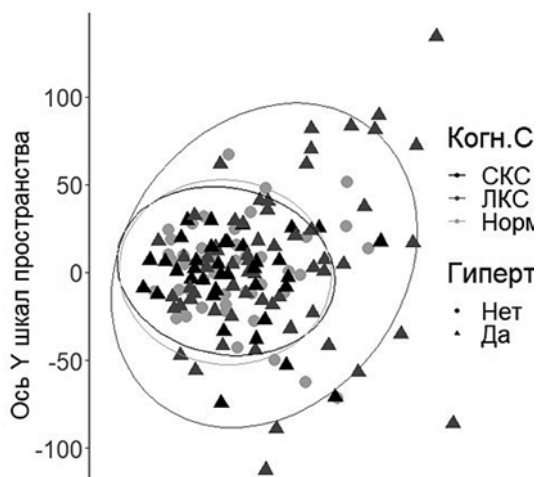


Рис. 1. Визуализация исходного пространства

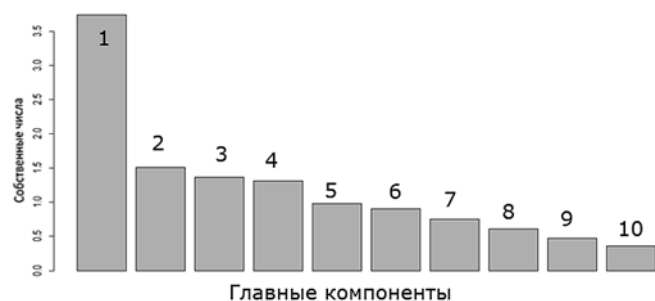


Рис. 2. Собственные числа главных компонент исходного пространства

На рис. 1 представлена визуализация исходного пространства с использованием метрического многомерного шкалирования в евклидовом пространстве [18]. Также данные были подвергнуты процедуре стандартизации.

На рис. 1 представлены разными цветами и формами маркеров пациенты, относящиеся к разным группам. Треугольными маркерами отмечены пациенты с АГ, круглыми — пациенты из группы контроля. Пациенты группы контроля также отмечены светлым цветом, темный цвет треугольного маркера соответствует пациентам с СКС, более светлый — с ЛКС. Из графика следует, что в исходном пространстве фактически отсутствует возможность алгоритмического разделения пациентов по группам по тяжести когнитивных нарушений.

Предположение о плохой согласованности данных в исходном виде подтверждается и результатами компонентного анализа. Использован метод главных компонент, являющийся эффективным для сокращения размерности пространства путем преобразования данных в новой системе координат, где каждая координата (компонента) максимизирует остаточную выборочную дисперсию. Так, исходя из анализа собственных чисел главных компонент (рис. 2) можно сделать вывод о том, что первые две главные компоненты объясняют меньше 50 % вариации исходного пространства, что говорит о несогласованности исходных данных и требует их преобразования в новое пространство за счет сокращения набора исходных признаков.

Кроме того, все факторные нагрузки 1-й и 2-й главных компонент меньше по модулю 0,5, что говорит об их слабой связанности с исходными признаками пространства, а следовательно, о некоей "бессистемности" распределения результатов оценки когнитивных функций.

Следовательно, необходимо преобразование исходных данных для поиска устойчивых комбинаций признаков, пригодных для алгоритмического разделения априорных групп пациентов. В рамках текущего пространства использование предиктивных алгоритмов будет неэффективным в силу, во первых, сильного переобучения при использовании в качестве зависимой переменной бинарных характеристик о принадлежности пациента к определенной группе, во вторых, отсутствия явных однофакторных зависимостей между признаками принадлежащих к разным группам, в третьих, существенно более сильного разделения пациентов на группы с АГ и контрольную, нежели на группы с наличием когнитивных нарушений и без него.

3. Описание алгоритма модификации исходного пространства данных

Для решения задачи по поиску оптимальных нейropsychологических шкал, позволяющих разделить группу пациентов с АГ на подгруппы пациентов с СКС и ЛКС, рассматривался вариант свертки исходного пространства и перехода к одному обобщенному признаку.

Для поиска подобного обобщающего показателя были исследованы все возможные комбинации психоневрологических параметров с использованием логистической регрессии (логит-модели) [19] для наилучшего решения следующих двух задач:

- разделение группы контроля и пациентов с когнитивными нарушениями;
- выделение внутри пациентов с КН групп с СКС и ЛКС.

Для каждого набора параметров проводили 1000 испытаний с варьированием пациентов между обучающей выборкой и тестовой выборкой. Выборки были разбиты по Парето в пропорции 80/20.

Для каждой модели анализировали следующие критерии валидации:

- коэффициент детерминации МакФаддена ($R_{McFadden}^2$) [20];
- информационный критерий Акаике (AIC) [21];
- число угаданных на тестовых выборках значений.

Для всех вариантов ресемплирования рассчитывались средние значения критериев верификации (медиана в случае числа угаданных).

По каждому критерию валидации отбирали 15 наилучших комбинаций оцениваемых когнитивных тестов. Среди всех комбинаций были отобраны те, что чаще всего попадали в 15 лучших по шести валидационным критериям. Также было рассчитана частота попаданий отдельного признака в наилучшие комбинации.

На основании этих результатов далее были исследованы четыре комбинации признаков. Вошедшие в каждую комбинацию признаки отмечены светлым цветом в табл. 2

Далее представлены результаты анализа по четвертой комбинации как оказавшейся наилучшей.

4. Компонентный анализ сокращенного пространства

На рис. 3 представлена визуализация пространства признаков, принадлежащих к четвертой комбинации признаков с использованием многомерного шкалирования. Цветовые обозначения и значения маркеров идентичны рис. 1.

Когнитивные признаки используемые для дальнейшего анализа в разных комбинациях

Тест	Комбинация 1	Комбинация 2	Комбинация 3	Комбинация 4
TMT A (с)	FALSE	TRUE	TRUE	FALSE
TMT B (с)	FALSE	TRUE	FALSE	TRUE
TMT B-TMT A (с)	TRUE	TRUE	TRUE	TRUE
MoCA тест (баллы)	FALSE	TRUE	TRUE	FALSE
Литеральные ассоциации (число слов)	FALSE	FALSE	FALSE	FALSE
Категориальные ассоциации (число слов)	FALSE	FALSE	FALSE	FALSE
Тест Струпа — T1 (с)	FALSE	FALSE	FALSE	TRUE
Тест Струпа — T2 (с)	FALSE	FALSE	FALSE	FALSE
Тест Струпа — T3 (с)	FALSE	FALSE </td <td>FALSE</td> <td>FALSE</td>	FALSE	FALSE
Тест Струпа — Коэффициент интерференции (T3—T2)*	FALSE	FALSE	FALSE	FALSE
Тест "10 слов", непосредственное воспроизведение (число слов)	FALSE	TRUE	TRUE	FALSE
Тест "10 слов", отсроченное воспроизведение (число слов)	TRUE	TRUE	TRUE	TRUE

* T3—T2 — разность затраченного времени на выполнение частей 2 и 3 теста Струпа

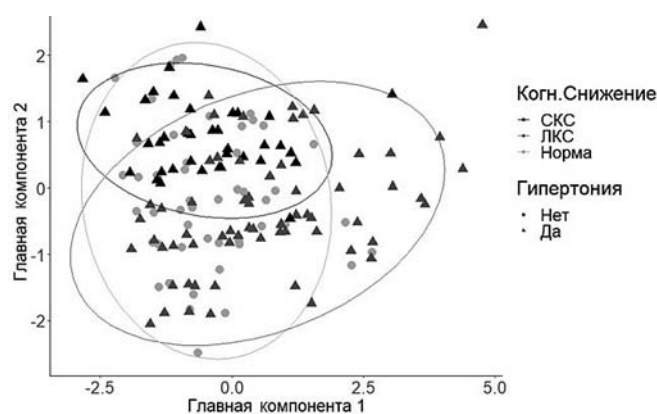


Рис. 3. Визуализация нового пространства

Из графика видно, что априорные группы стали разнесены в пространстве значительно лучше, нежели на рис. 1, хотя общее качество разделения не идеально. Для данного пространства коэффициент конкордации Кенделла составил $W = 0,4107$ при $p\text{-value} = 1,069 \cdot 10^{-06}$, таким образом удалось в несколько раз увеличить согласованность данных. Забегая вперед, отметим, что визуализация методом главных компонент будет полностью совпадать с многомерным шкалированием. В целом группа

Таблица 3

Факторные нагрузки главных компонент нового пространства

Тест	Главная компонента 1	Главная компонента 2	Главная компонента 3	Главная компонента 4
ТМТ В (с)	0,661		0,263	0,703
ТМТ В-ТМТ А (с)	0,664		0,222	-0,709
Тест Струпа — Т1 (с)	0,33	0,146	-0,932	
Тест "10 слов", отсроченное воспроизведение (число слов)	-0,117	0,985	0,114	

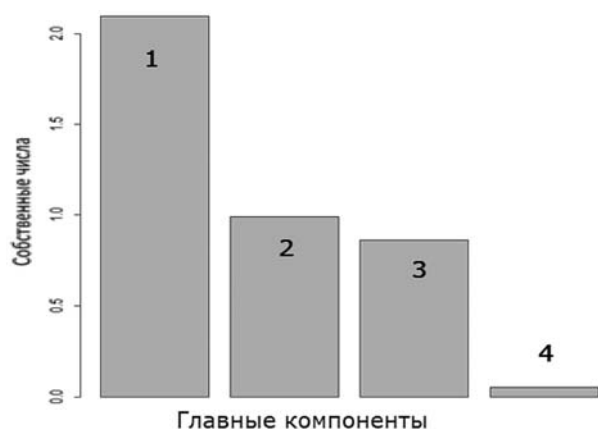


Рис. 4. Собственные числа главных компонент нового пространства

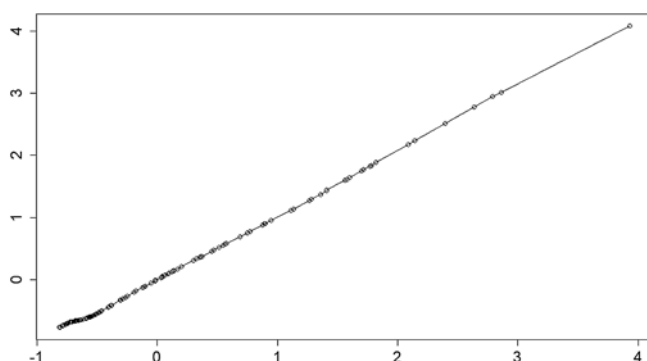


Рис. 5. Главная кривая нового пространства

контроля и пациенты с СКС обладают более низкими значениями по 1-й главной компоненте, в то время как пациенты с СКС визуально явно отличаются от пациентов с ЛКС по значениям 2-й компоненты.

Компонентный анализ показывает лучшую согласованность данных нового пространства как с точки зрения *собственных чисел* (рис. 4), так и с точки зрения *факторных нагрузок* (см. табл. 3, где приведены значения только значимых нагрузок).

Стоит отметить, что первая компонента (разделяющая группу контроля и СКС от ЛКС) имеет большую связанность с тестами связи цифр и букв, в то время как 2-я компонента, лучше отделяющая группу с субъективными нарушениями от группы с легкими нарушениями, связана с тестом "10 слов".

Также для проверки возможностей нелинейного разложения пространства данных был применен метод главной кривой, который заключается в нахождении гладкой кривой, которая проходит через середину данных в ортогональном смысле. Эта кривая является непараметрическим обобщением линейного метода главных компонент и характеризует основной характер зависимости [22]. В нашем случае видно (рис. 5), что кривая, по сути, линейна, что говорит об отсутствии значимой нелинейной связи в новом пространстве. Соответственно, можно использовать полученные ранее главные компоненты для дальнейшего анализа.

5. Дискриминантный анализ в новом пространстве

Была исследована возможность применять полученные главные компоненты для предиктивных и регрессионных моделей. Был применен стандартный метод MANOVA [23] и непараметрический тест Краскела— Уоллиса [24] для проверки гипотезы о способности полученных главных компонент эффективно разделять пациентов по априорным группам.

Таблица 4

Результаты многомерного дисперсионного анализа

Главные компоненты		Число степеней свободы	Сумма квадратов ошибки	Средняя ошибка	Значение F-критерия	P-value
1-я главная компонента	Группа пациентов	2	40,308	20,1539	10,796	$4,18 \cdot 10^{-5}$
	Остатки	149	278,14	1,8667		
2-я главная компонента	Группа пациентов	2	29,499	14,7496	18,203	$8,46 \cdot 10^{-8}$
	Остатки	149	120,732	0,8103		

Результаты многомерного теста (табл. 4) являются значимыми, соответственно, можем предположить, что найденные компоненты являются эффективными с точки зрения разделения групп пациентов.

Метод MANOVA применяется только для нормально распределенных данных, однако применение теста Шапиро—Уилкса показало, что нормально распределена только 1-я компонента. Поэтому для уточнения результатов анализа был проведен тест Краскела—Уоллиса. Для 1-й компоненты было получено значение критерия $\chi^2 = 16,26$ при $p\text{-value} = 0,0002945$, для 1-й компоненты было получено значение критерия $\chi^2 = 33,248$ при $p\text{-value} = 6,031 \cdot 10^{-8}$, что подтверждает выводы по методу MANOVA.

Заключение

Поиск и обнаружение факторов, способных влиять на когнитивные нарушения на ранних этапах их проявления, являются важной статистической и алгоритмической медицинской и социальной задачей. Однако в силу сложности этих процессов классические методы регрессионо-корреляционного анализа оказываются неприменимы. Более того, как показывает данное исследование, выбор в качестве зависимых переменных априорных принадлежностей к той или иной группе не позволит связать физиологические отклонения с когнитивными расстройствами. Использование свертки пространства результатов нейропсихологического тестирования пациентов для перехода к интегральным характеристикам применяется в предикторных и регрессионных моделях как отклик системы.

Предложенная в статье методика с применением современных инструментальных средств для обработки статистических данных, таких как язык R, позволяет в несколько раз повысить согласованность итогового пространства интегральных характеристик. Также было показано, что полученные главные компоненты обладают статистической связью с пространством физиологических показателей силой, и эта методика может быть применима и для аналогичных исследований, связанных с разнородностью и сложными правилами группировки исходных количественных данных.

Информационная система поддержки принятия решений врачом, построенная на базе созданной методики, выявленных факторов и разработанных алгоритмов, может найти широкое применение в практике медицинских учреждений при диагностики когнитивных нарушений.

Список литературы

1. DerSimonian R., Kacker R. Random-effects model for meta-analysis of clinical trials: An update // *Contemporary Clinical Trials*. 2007. Vol. 28, Iss. 2. P. 105—114.
2. Paule R. C., Mandel J. Consensus values and weighting factors // *Journal of Research of the National Bureau of Standards*. 1982. Vol. 87 (5). P. 377—385.
3. Lee K. J., Thompson S. G. Flexible parametric models for random-effects distributions // *Statistics in Medicine*. 2008. Vol. 27 (3). P. 418—434.
4. Baker R., Jackson D. A new approach to outliers in meta-analysis // *Health Care Management Science*. 2008. Vol. 11 (2). P. 121—131
5. Аронов П. М. Оценка согласованного значения результатов межлабораторных измерений с минимальным увеличением их неопределенности // *Стандартные образцы*. 2019. № 4.
6. Красавчиков В. О. Математические методы, алгоритмы и технология геомоделирования для решения задач геологии нефти и газа. Автореф. дис. ... док. техн. наук. Кемерово, 2007. 39 с.
7. Яхно, Н. Н., Захаров В. В., Коберская Н. Н., Мхитарян Э. А., Гришина Д. А., Локшина А. Б., Савушкина И. Ю., Посохов С. И. "Предумеренные" (субъективные и легкие) когнитивные расстройства // *Неврологический журнал*. 2017. Т. 22, № 4. С. 198—204.
8. Reisberg B., Shulman M. B., Torossian C., Leng L., Zhu W. Outcome over seven years of healthy adults with and without subjective cognitive impairment // *Alzheimers Dement*. 2010. Vol. 6. P. 11—24.
9. Чазова И. Е. и др. Клинические рекомендации. Диагностика и лечение артериальной гипертензии // *Кардиологический вестник*. 2015. № 1. С. 3—30.
10. Парфенов В. А., Остроумова Т. М., Перепелова Е. М., Перепелов В. А., Кочетков А. И., Остроумова О. Д. Перфузия головного мозга, когнитивные функции и сосудистый возраст у пациентов среднего возраста с эссенциальной артериальной гипертензией // *Кардиология*. 2018. Т. 58, № 5. С. 23—31.
11. Остроумова Т. М., Парфенов В. А., Остроумова О. Д., Перепелова Е. М., Перепелов В. А., Борисова Е. В. Возможности метода бесконтрастной магнитно-резонансной перфузии для выявления раннего поражения головного мозга при эссенциальной артериальной гипертензии // *Неврология, нейропсихиатрия, психосоматика*. 2018. Т. 10, № 1. С. 17—23.
12. Лурья А. Р. Заучивание 10 слов // *Альманах психологических тестов*. 1995. С. 92—94.
13. Nasreddine Z. S. et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment // *Journal of the American Geriatrics Society*. 2005. Vol. 53 (4).
14. Nasreddine Z. S. MoCA Russian version, 2010. Available from: http://www.mocatest.org/wp-content/uploads/2015/testsinstructions/MoCA-Test-Russian_2010.Pdf
15. Reitan R. Validity of the Trail Making Test as an indicator of organic brain damage // *Percept Mot Skills*. 1958. Vol. 8. P. 271—276.
16. MacLeod C. M. Half a century of research on the Stroop effect: An integrative review // *Psychol Bull*. 1991. Vol. 109, N. 2. P. 163—203.
17. Парфенов В. А., Захаров В. В., Преображенская И. С. Когнитивные расстройства. М.: Ремедиум, 2014. 192 с.
18. Mardia K. V. Some properties of classical multidimensional scaling. *Communications on Statistics // Theory and Methods*. Vol. 7. P. 1233—1241. doi: 10.1080/03610927808827707.
19. Магнус Я. Р., Катышев П. К., Пересецкий А. А. М13 Эконометрика. Начальный курс: Учеб. 6-е изд., перераб. и доп. М.: Дело, 2004. 576 с.
20. Greene W. H. *Econometric Analysis*. Prentice Hall, Upper Saddle River, 2003.
21. Akaike H. A new look at the statistical model identification // *IEEE Transactions on Automatic Control*. 1974. Vol. 19. P. 716—723.
22. Hastie T., Stuetzle W. Principal Curves // *JASA*. Jun., 1989. Vol. 84, N. 406 P. 502—516.
23. Krzanowski W. J. *Principles of Multivariate Analysis. A User's Perspective*. Oxford, 1988.
24. Hollander M., Wolfe D. A. *Nonparametric Statistical Methods*. New York: John Wiley & Sons, 1973. P. 115—120.

N. N. Yakhno^{1, 2}, Professor, e-mail: info@ditc.ras.ru, V. N. Gridin¹, Professor, e-mail: info@ditc.ras.ru,
D. S. Smirnof¹, Senior Researcher, e-mail: info@ditc.ras.ru,
V. S. Panishchev¹, Senior Researcher, e-mail: info@ditc.ras.ru,
V. A. Parfenov², Professor, e-mail: info@ditc.ras.ru,
T. M. Ostroumova^{1, 2}, Junior Researcher, e-mail: info@ditc.ras.ru,
N. N. Koberskaya^{1, 2}, Senior Researcher, e-mail: info@ditc.ras.ru,

¹ Design Information Technologies Center Russian Academy of Sciences,
Odintsovo, Moscow Region, Russian Federation

² I. M. Sechenov First Medical State University, Moscow

Statistical Processing and Methods for Reducing the Dimension of Space on the Example of Data for Patients in the Analysis of Cognitive Impairment

The article proposes the development of methods for statistical processing of information and algorithms for converting and analyzing medical data of patients in order to detect factors of subjective and easy cognitive decline. The inapplicability of classical classification models for such data with a low initial consistency is shown. A technique is proposed for identifying groups of indicators that have the greatest consistency and dividing ability for a priori groups of patients. The data transformed in accordance with the methodology can be the initial data for constructing effective predictive models of the regression type.

Keywords: cognitive decline, spatial dimensionality reduction, multidimensional statistical methods, logistic regression, principal component analysis

Acknowledgements: The reported study was funded by RFBR, project number № 19-29-01112 mk

DOI: 10.17587/it.26.515-522

References

1. DerSimonian R., Kacker R. Random-effects model for meta-analysis of clinical trials: An update, *Contemporary Clinical Trials*, 2007, vol. 28, iss. 2, pp. 105–114.
2. Paule R. C., Mandel, J. Consensus values and weighting factors, *Journal of Research of the National Bureau of Standards*, 1982, vol. 87 (5), pp. 377–385.
3. Lee K. J., Thompson S. G. Flexible parametric models for random-effects distributions, *Statistics in Medicine*, 2008, vol. 27 (3), pp. 418–434.
4. Baker R., Jackson D. A new approach to outliers in meta-analysis, *Health Care Management Science*, 2008, vol. 11 (2), pp. 121–131.
5. Aronov P. M. Estimation of the agreed value of the results of inter-laboratory measurements with a minimum increase in their uncertainty, *Standard Samples*, 2019, no. 4 (in Russian).
6. Krasavchikov V. O. Mathematical methods, algorithms and geomodeling technology for solving problems of oil and gas geology, Abstract of a dissertation of a doctor of technical science, Kemerovo, 2007, 39 p. (in Russian).
7. Yakhno N. N., Zakharov V. V., Koberskaya N. N., Mkhitarian E. A., Grishina D. A., Lokshina A. B., Savushkina I. Yu., Posokhov S. I. Premeditated (subjective and mild) cognitive, *Neurological Journal*, 2017, vol. 22, no. 4, pp. 198–204 (in Russian).
8. Reisberg B., Shulman M. B., Torossian C., Leng L., Zhu W. Outcome over seven years of healthy adults with and without subjective cognitive impairment, *Alzheimers Dement.*, 2010, vol. 6, pp. 11–24.
9. Chazova I. E. et al. Clinical recommendations. Diagnosis and treatment of arterial hypertension, *Cardiological Bulletin*, 2015, no. 1, pp. 3–30 (in Russian).
10. Parfenov V. A., Ostroumova T. M., Perepelova Ye. M., Perepelov V. A., Kochetkov A. I., Ostroumova O. D. Brain perfusion, cognitive functions and vascular age in middle-aged patients with essential arterial hypertension, *Cardiology*, 2018, vol. 58, no. 5, pp. 23–31 (in Russian).
11. Ostroumova T. M., Parfenov V. A., Ostroumova O. D., Perepelova Ye. M., Perepelov V. A., Borisova Ye. V. Possibilities of the method of non-contrast magnetic resonance perfusion for detecting early brain damage in case of essential arterial hypertension, *Neurology, Neuropsychiatry, Psychosomatics*, 2018, vol. 10, no. 1, pp. 17–23 (in Russian).
12. Luria A. R. Memorization of 10 words, *Almanac of Psychological Tests*, 1995, pp. 92–94 (in Russian).
13. Nasreddine Z. S., et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment, *Journal of the American Geriatrics Society*, 2005, vol. 53 (4).
14. Nasreddine Z. S. MoCA Russian version, 2010, available at: http://www.mocatest.org/wp-content/uploads/2015/testsinstructions/MoCA-Test-Russian_2010.pdf.
15. Reitan R. Validity of the Trail Making Test as an indicator of organic brain damage, *Percept Mot Skills*, 1958; 8:271–276.
16. MacLeod C. M. Half a century of research on the Stroop effect: An integrative review, *Psychol Bull* 1991; 109 (2): 163–203.
17. Parfenov V. A., Zakharov V. V., Preobrazhenskaya I. S. Cognitive disorders. Moscow, Remedium, 2014, 192 p. (in Russian).
18. Mardia K. V. Some properties of classical multidimensional scaling, *Communications on Statistics Theory and Methods*, vol. 7, pp. 1233–1241. doi: 10.1080/03610927808827707.
19. Magnus Ya. R., Katyshev P. K., Peresetskiy A. A. M13 Econometrics. Initial course: Textbook, Moscow, Business, 2004, 576 p. (in Russian).
20. Greene W. H. Econometric Analysis. 5th Edition, Prentice Hall, Upper Saddle River, 2003.
21. Akaike H. A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 1974, vol. 19, pp. 716–723.
22. Hastie T., Stuetzle W. Principal Curves, *JASA*, 1989, vol. 84, no. 406, pp. 502–516.
23. Krzanowski W. J. Principles of Multivariate Analysis. A User's Perspective, Oxford, 1988.
24. Hollander M., Wolfe D. A. Nonparametric Statistical Methods, New York, John Wiley & Sons, 1973, pp. 115–120.