

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ: ОПЫТ ИСПОЛЬЗОВАНИЯ ПАКЕТА PSPP В МЕДИЦИНСКОМ ВУЗЕ ДЛЯ ФОРМИРОВАНИЯ ПРОФЕССИОНАЛЬНЫХ КОМПЕТЕНЦИЙ

INFORMATION TECHNOLOGIES IN STATISTICAL ANALYSIS: EXPERIENCE OF USAGE OF PACKAGE PSPP IN MEDICAL SCHOOL FOR FORMING PROFESSIONAL COMPETENCIES

**S. Kasyuk
T. Shamaeva**

Summary. The article describes the information technologies used to form the professional competencies in the field of statistical analysis of medical information on the basis of evidence-based medicine. It is shown that using of statistical package PSPP provide the basic capabilities to analyze medical information. The method of forming these professional competencies to studying the discipline of medical informatics is described.

Keywords: information technologies, statistical analysis, package PSPP, clinical studies, evidence-based medicine, professional competencies.

Касюк Сергей Тимурович

*К.т.н., доцент, ФГБОУ ВО «Южно-Уральский
государственный медицинский университет»
Министерства здравоохранения Российской Федерации
(г. Челябинск)
sergey.kasyuk@gmail.com*

Шамаева Татьяна Николаевна

*К.п.н., доцент, ФГБОУ ВО «Южно-Уральский
государственный медицинский университет»
Министерства здравоохранения Российской Федерации
(г. Челябинск)
shamtan@rambler.ru*

Аннотация. Статья посвящена информационным технологиям, используемым для формирования профессиональных компетенций в области статистического анализа медицинской информации на основе доказательной медицины. Показано, что использование статистического пакета PSPP обеспечивает основные возможности для статистического анализа медицинской информации. Приведена методика формирования указанных профессиональных компетенций при изучении дисциплины «Медицинская информатика».

Ключевые слова: информационные технологии, статистический анализ, пакет PSPP, клинические исследования, доказательная медицина, профессиональные компетенции.

Профессиональная деятельность будущих выпускников медицинского вуза будет строиться с учетом как *статистического анализа медицинской информации о работе лечебного учреждения* и публичного представления полученных результатов, так и *статистического анализа информации при решении научно-исследовательских и научно-прикладных задач в области здравоохранения.*

В ФГОС третьего поколения предусмотрено формирование *профессиональных компетенций* обучающихся, связанных со *статистическим анализом медицинской информации и использованием методов доказательной медицины*, которые в настоящее время реализуется с применением соответствующих информационных аналитических технологий. Данные компетенции являются обязательными для обучающихся.

Согласно требованиям ФГОС ВО по направлению подготовки 31.05.01 «Лечебное дело», утверждено-

го приказом № 95 Минобрнауки России от 09.02.2016, в результате освоения программы специалитета выпускник должен обладать рядом *профессиональных компетенций (ПК) в медицинской деятельности и научно-исследовательской деятельности*, связанных со *статистическим анализом медицинской информации*. К таким компетенциям следует отнести: способность и готовность к применению социально-гигиенических методик сбора и микро-статистического анализа информации о показателях здоровья населения (ПК-4), готовность к анализу и публичному представлению медицинской информации на основе *доказательной медицины* (ПК-20) [1].

Таким образом, необходимо на уровне высшего профессионального образования сформировать у обучающихся *компетенции* как в областях *статистического анализа медицинской информации*, так и использования *соответствующих статистических пакетов.*

Современный *статистический анализ медицинской информации* базируется на использовании таких статистических пакетов, как SPSS компании IBM, STATISTICA компании TIBCO Software Inc., SAS компании SAS Institute Inc., свободного пакета PSPP, а также статистического языка R.

Статья ставит *целью* изучить опыт и проанализировать перспективы использования пакета PSPP для формирования *профессиональных компетенций в области статистического анализа медицинской информации на основе доказательной медицины* у обучающихся по направлению подготовки 31.05.01 «Лечебное дело» в Южно-Уральском государственном медицинском университете (ЮУГМУ).

Профессиональные компетенции в рабочих программах и планах

Обучающиеся в ЮУГМУ по направлению подготовки 31.05.01 «Лечебное дело» изучают на первом и втором курсах две дисциплины, связанные с медико-статистическим анализом информации: «Физика, математика» и «Медицинская информатика». Эти дисциплины относятся к базовой части программы специалитета.

Согласно рабочих программ дисциплин «Физика, математика», «Медицинская информатика», принятых в ЮУГМУ, в результате изучения этих дисциплин обучающиеся должны:

- ◆ знать теоретические *основы статистического анализа информации*, принципы, методы *математической статистики и доказательной медицины*;
- ◆ уметь использовать *базовые методы статистики* при обработке и представлении экспериментальных данных, *анализировать и представлять медицинскую информацию* в соответствии с принципами математической статистики;
- ◆ владеть методами *анализа статистической информации* о показателях здоровья населения, *технологией анализа биомедицинских данных с позиций математической статистики*.

При изучении дисциплины «Физика, математика» обучающиеся знакомятся с основными понятиями математической статистики и овладевают приемами расчёта числовых характеристик выборок, а также точечных и интервальных оценок параметров генеральной совокупности. На занятиях по дисциплине «Медицинская информатика» происходит дальнейшее *обогащение понятий математической статистики*, формируются *навыки работы в статистическом пакете*. Так, в учебном плане дисциплины «Медицинская информа-

тика» [2], принятом в ЮУГМУ, присутствуют темы № 12–14 «Математическая обработка медико-биологических данных с помощью статистического пакета».

В ЮУГМУ в качестве средства формирования компетенций, связанных со *статистическим анализом медицинской информации*, был выбран свободный статистический пакет PSPP версии 0.10.2.

Учебный процесс дисциплин «Физика, математика», «Медицинская информатика» реализуется как целостная система организационных форм и методов обучения (лекций, практических занятий и самостоятельной работы), образующих дидактический цикл обучения.

Доказательная медицина и её роль в клинических исследованиях

Доказательная медицина — это использование результатов лучших клинических исследований для выбора лечения конкретного пациента, это интеграция лучших научных доказательств с клиническим опытом и ожиданиями пациентов [3].

Р. Флетчер, С. Флетчер, Э. Вагнер в своей классической работе [4] пишут, что «термин «доказательная медицина» был введен для обозначения медицинской практики, основанной на данных хорошо организованных клинических исследований. Задача врача — найти эти данные..., оценить их, всесторонне обдумать и применить к своему пациенту».

Т. Гринхальх постулирует, что «...в основе доказательной медицины лежит *использование количественных сведений*, полученных в исследованиях на группах пациентов, при принятии решений о конкретных пациентах» [5].

Клиническая наука особенно убедительна, когда может обеспечить *количественный подход при измерениях*. Отчасти это обусловлено тем, что *количественные результаты* дают более достоверные доказательства, позволяют оценить ошибку и облегчают обмен информацией как между врачами, так и между врачом и больным [4].

Принципы *доказательной медицины*, широко используемые при проведении *клинических исследований*, позволяют дать объективную сравнительную оценку эффективности и безопасности применения изучаемого метода или вмешательства для лечения определенных патологических состояний [3].

Значение *клинических исследований* заключается в поиске эффективных и безопасных методов диагности-

ки, лечения и профилактики и, на этой основе, создания рекомендаций для широкого применения [3].

К основным этапам клинических исследований следует отнести: формулировку целей и выдвижение гипотез, планирование и выполнение исследований, сбор и подготовку данных, анализ данных, интерпретацию результатов, представление результатов, формулировку выводов [6].

Современные информационные аналитические технологии позволяют решать задачи *основных этапов клинических исследований* [7, 8]:

1. Планирование медицинских исследований.
2. Рандомизация пациентов.
3. Вычисление необходимого объема выборки, анализ мощности.
4. Сбор и подготовка данных для анализа: консолидация, агрегирование, трансформация, группировка, очистка.
5. Вычисление основных описательных статистик выборок.
6. Визуализация данных.
7. Установление степени связи между различными категориальными данными. Построение таблиц сопряженности. Вычисление абсолютных, относительных рисков, отношения шансов.
8. Выявление зависимостей между показателями пациентов: вычисление коэффициентов корреляции Пирсона, Спирмена, Тау Кендалла.
9. Выявление эффективности лечения в нескольких выборках пациентов: параметрические критерии (дисперсионный анализ Фишера, t -критерии) и непараметрические критерии (U -критерий Манна-Уитни, H -критерий Краскела — Уоллиса, дисперсионного анализ Фридмана, критерий Колмогорова — Смирнова, критерий Вилкоксона, критерий знаков).
10. Прогнозирование исхода лечения и выявление факторов, влияющих на этот исход: дискриминантный анализ, деревья классификации, линейная регрессия, логистическая регрессия.
11. Анализ выживаемости, сравнение выживаемости в различных выборках: критерии Гехана — Вилкоксона, Кокса — Ментела, F -критерий Кокса, логарифмический ранговый критерий.
12. Представление результатов и оформление отчетов.

Статистический пакет PSPP

Применяемый в ЮУГМУ пакет PSPP является свободным программным обеспечением для статистического анализа данных в ОС Windows, распространяемым под лицензией GNU (General Public License). PSPP является

свободной заменой коммерческого пакета SPSS компании IBM со схожими функциональными возможностями [9].

В PSPP реализованы следующие статистические процедуры [10, 11]:

1. *Вычисление описательных статистик.*
2. *Сравнение средних: t -критерий для одной выборки, t -критерий для независимых выборок, t -критерий для зависимых выборок, однофакторный дисперсионный анализ (ANOVA).*
3. *Параметрическая корреляция Пирсона.*
4. *Кластерный анализ методом k -средних.*
5. *Факторный анализ.*
6. *Анализ надежности.*
7. *Регрессионный анализ: линейный и логистический.*
8. *Непараметрическая статистика: сравнения частот в двух независимых выборках по критерию Хи-квадрат, проверка закона распределения выборок по критерию Колмагорова — Смирнова, сравнение двух зависимых выборок с помощью непараметрического критерия Вилкоксона, сравнение двух зависимых выборок с помощью непараметрического знакового критерия, сравнение частот в двух зависимых выборках по критерию Мак-Немара, критерий Q Кохрена для повторных испытаний.*
9. *Кривая операционных характеристик (ROC-кривая).*
10. *Визуализация данных: построение графиков, гистограмм.*

Методика формирования профессиональных компетенций

За основу *методики формирования профессиональных компетенций* в области *статистического анализа медицинской информации с использованием пакета PSPP* взята методическая система доцентов Т.Н. Шамаевой, О.А. Степановой, описанная в их работе по научно-методическому обеспечению дисциплины «Информатика и медицинская статистика» [12].

На кафедре *математики, медицинской информатики, информатики и статистики, физики ЮУГМУ* были разработаны соответствующие лекции, практические работы и фонды оценочных средств.

На лекциях даётся представление о принципах доказательной медицины, этапах клинических исследований, базовых понятиях статистического анализа (генеральной совокупности, выборке, основных статистиках), видах статистических гипотез, параметрических и непараметрических критериях и различных видах анализа.

Таблица 1

Переменная	N	Среднее	Станд. отклон.	Дисперсия	Станд. ошибка средн. знач.	Экссесс	Асимметрия
Плацебо	11	2,26	0,31	0,10	0,09	-1,12	0,15
Нифедипин	11	2,08	0,41	0,17	0,12	-1,11	-0,46

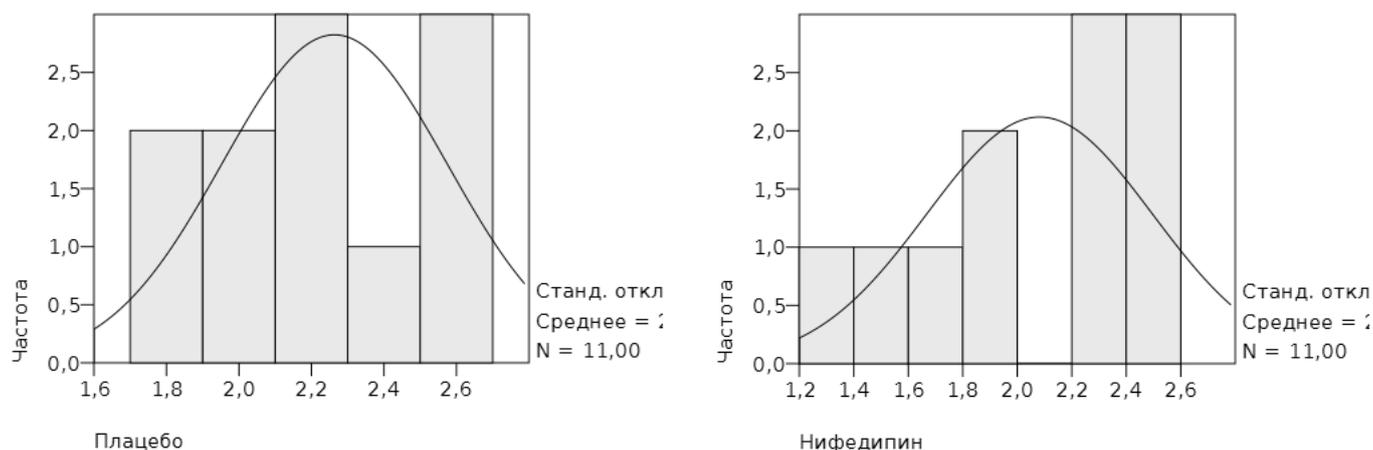


Рис 1. Гистограммы переменных Плацебо и Нифедипин

В содержание одной из лекций включен обзор интерфейса PSPP, подготовка и работа с данными в этом пакете.

Во время практических занятий обучающиеся, используя пакет PSPP, для каждой решаемой задачи: определяют тип и структуру исходных медицинских данных, вводят информацию в таблицы пакета, вычисляют описательные статистики, определяют вид распределений, формулируют нулевую и альтернативную гипотезы в зависимости от цели анализа, вычисляют статистики критерия, делают статистически обоснованные выводы.

Практика формирования профессиональных компетенций

Проиллюстрируем практику формирования компетенций в пакете PSPP на учебной задаче сравнения двух независимых выборок с помощью t -критерия. Дана информация о влиянии диуретиков на диаметр коронарных сосудов в двух независимых выборках пациентов по 11 человек, принимавших плацебо и нифедипин. Необходимо определить влияние нифедипина на диаметр коронарных сосудов [7].

Статистический анализ медицинской информации в пакете PSPP для этой задачи проводится по следующему алгоритму:

1. Создается таблица наблюдений, состоящей из столбцов (переменных *Плацебо* и *Нифедипин*)

тин) и строк (наблюдений). Вводятся исходные данные, соответствующие двум выборкам — *Плацебо* и *Нифедипин*.

2. Вычисляются описательные статистики выборок (табл. 1).
3. Проверяется нулевая гипотеза H_0 о том, что выборки извлечены из генеральных совокупностей с нормальным распределением изучаемых признаков. Проверка осуществляется по критерию Колмогорова — Смирнова и выборочным коэффициентам асимметрии и эксцесса. Оказывается, что расчетные значения критерия Колмогорова — Смирнова (0,13 для *Плацебо* и 0,16 для *Нифедипин*) для выборок меньше табличного значения 0,39 для заданного количества наблюдений $n = 11$ и уровня значимости $\alpha = 0,05$, и, следовательно, нулевая гипотеза H_0 о нормальности распределения выборок *Плацебо* и *Нифедипин* подтверждается.
4. Строятся гистограммы для исследуемых выборок *Плацебо* и *Нифедипин* (рис. 1). При небольшом количестве наблюдений эти гистограммы отличаются по внешнему виду от гистограмм для выборок с нормальным распределением (см. таблицу).
5. Убедившись, что распределение в выборках нормальное, данные перестраиваются в соответствии с требованием PSPP. В первый столбец, называемый **Диаметр**, вводятся величины диаметра коронарных сосудов. Вторым столбцом

Набл.	Диаметр	Независимая
1	2,2	плацебо
2	2,2	плацебо
3	2,6	плацебо
4	2,0	плацебо
5	2,1	плацебо
6	1,8	плацебо
7	2,4	плацебо
8	2,3	плацебо
9	2,7	плацебо
10	2,7	плацебо
11	1,9	плацебо
12	2,5	нифедипин
13	1,7	нифедипин
14	1,5	нифедипин
15	2,5	нифедипин
16	1,4	нифедипин
17	1,9	нифедипин
18	2,3	нифедипин
19	2,0	нифедипин
20	2,6	нифедипин
21	2,3	нифедипин
22	2,2	нифедипин

Рис. 2

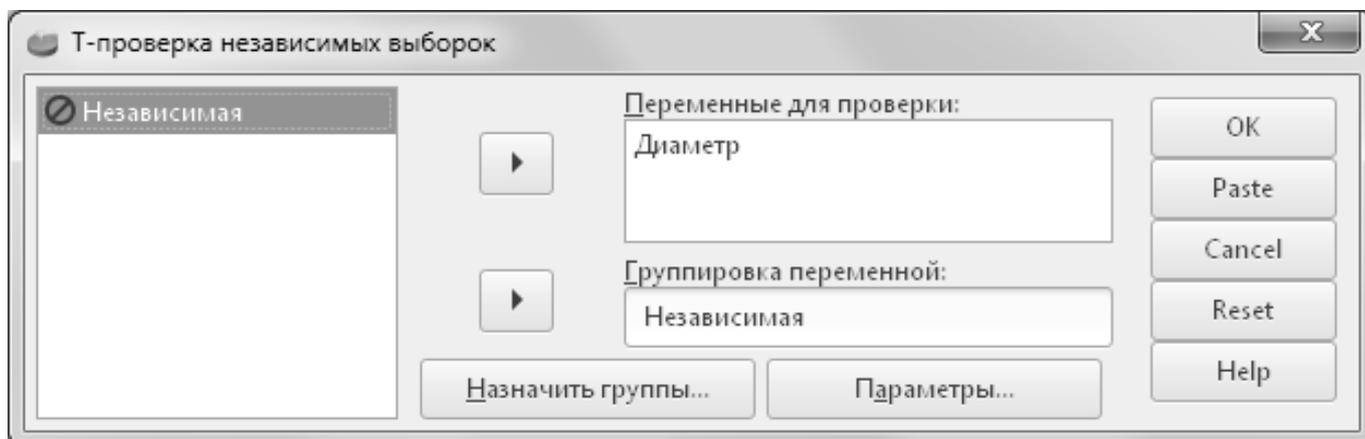


Рис. 3. Окно T-проверка независимых выборок

называемый **Независимая**, заполняется категориальными данными — именами переменных «плацебо» и «нифедипин» (рис. 2).

6. В окне **T-проверка независимых выборок** (рис. 3) выбирается переменная для проверки — **Диаметр** и группирующая переменная — **Независимая**. Назначаются для анализа две независимые группы — *плацебо* и *нифедипин*. Нажимается **ОК**.
7. В результате на экране появится таблица с результатами проверки независимых выборок по *t*-критерию (табл. 2) со следующими показателями:

- а) **F** — значение критерия Левина $F(1, df)$, где *df* — степень свободы;
- б) **Знач.** — уровень значимости *p* для критерия Левина;
- в) **t** — значение *t*-критерия для независимых выборок;
- г) **df** — число степеней свободы;
- д) **Знач. (двустороннее)** — уровень значимости для *t*-критерия;
- е) **Средняя разница** — разница средних значений выборок;

Таблица 2. Результаты проверка независимых выборок по t-критерию

Левина F	Левина Знач.	t	df	Знач. (двустороннее)	Средняя разница	Разница стандартных погрешностей	95%-ый доверительный интервал разницы	
							Нижняя	Верхняя
1,53	0,231	1,16	20	0,258	0,18	0,16	-0,14	0,51

ж) Разница стандартных погрешностей;

з) 95%-й доверительный интервал разницы.

- Проверяется нулевая гипотеза H_0 о равенстве дисперсий в выборках *Плацебо* и *Нифедипин* по критерию Левина. Поскольку расчетный уровень значимости для этого критерия $p = 0,231$ больше критического значения 0,05, то принимается гипотеза H_0 .
- Проверяется нулевая гипотеза H_0 об отсутствии влияния диуретиков на диаметр коронарных сосудов. Расчетный уровень значимости t -критерия $p = 0,258$ больше критического значения 0,05, а расчетное значение $t = 1,16$ меньше табличного значения 2,09 при числе степеней свободы $df = 20$ и уровне значимости $\alpha = 0,05$. Следовательно, принимается нулевая гипотеза H_0 .

Приведенная задача наглядно демонстрирует основную проблему: пакет PSPP рассчитывает в окне вывода десятки статистических показателей, смысл и значения которых должны быть понятны обучающимся.

Изучение практики формирования компетенций выявило недостаточный уровень подготовки обучающихся по математической статистике, приводящий к сложностям в интерпретации полученных результатов анализа в PSPP. Эту проблему можно решить за счет увеличения интенсивности учебного процесса, лучшей организации самостоятельной работы и проведения дополнительных консультаций.

Непосредственное применение пакета PSPP не представляет серьезных проблем, является эффективным для обучения, а также вызывает понимание и живой интерес со стороны обучающихся. Практически анализ данных сводится к использованию готовых статистических процедур, заложенных в PSPP, применению стандартных пошаговых статистических алгоритмов и интерпретации полученных результатов.

К достоинствам пакета статистического PSPP, обуславливающим перспективу его использования в высшей школе, следует отнести: свободное использование пакета PSPP как альтернативы SPSS; функциональные возможности пакета PSPP, достаточные для базового статистического анализа медицинской информации;

графический пользовательский интерфейс пакета PSPP с поддержкой русского языка.

К недостаткам использования статистического пакета PSPP следует отнести:

- ограниченные функциональные возможности по сравнению с такими коммерческими пакетами, как SPSS или STATISTICA, и отсутствие многих статистических критериев анализа информации;
- отсутствие отечественной литературы, посвященной пакету PSPP. Данный недостаток компенсируется литературой по пакету SPSS (авторы А.Н. Герасимов, П.Ю. Дубнов, А.О. Крыштановский, А.А. Макаров, А.Д. Наследов, Д.Н. Таганов), а также многочисленной литературой по обработке медико-биологических данных в пакете STATISTICA (авторы В.П. Боровиков, Э.А. Вуколов, О.Ю. Реброва, Н.В. Трухачёва, А.А. Халафян, В.И. Юнкеров). Среди англоязычных источников можно выделить авторов С. Halter и E. Bigham [10, 11];
- программные ошибки, приводящие к сбоям в работе пакета PSPP.

Проверка уровня сформированности профессиональных компетенций в области статистического анализа медицинской информации в пакете PSPP, осуществляющаяся в ЮУГМУ в виде тестирования в программе tTester, показывает хорошие результаты у обучающихся.

Перспективы формирования профессиональных компетенций

Формирование профессиональных компетенций в области статистического анализа медицинской информации позволит обучающимся на 6-м курсе грамотно выполнить НИР, предусмотренную ФГОС [1] в рамках учебной практики и относящуюся к базовой части программы специалитета, и в дальнейшем, в профессиональной деятельности, осуществлять медико-статистический анализ информации на основе доказательной медицины и публично представлять полученные результаты. В случае необходимости появляется возможность быстрого перехода на коммерческие

статистические пакеты с расширенными функциональными возможностями.

Как отмечают в своей статье Н. И. Брико, В. И. Покровский, «в области образования *доказательная медицина* меняет существо как до-, так и последипломного медицинского образования. Сегодня в развитых странах врачи изучают не столько стандартные курсы, сколько самообучаются, разыскивая надежные ответы на самые важные вопросы собственной практики, т.е. приоритет отдается самостоятельной работе. Система непрерывного медицинского образования, завоевавшая сегодня наибольшее признание во всем мире, во многом основана на использовании современных информационных технологий» [13]. В перспективе клинические исследования и медицинское образование в аспекте доказатель-

ности принимаемых решений должны интегрироваться в единую систему [14].

ВЫВОДЫ

В работе представлен опыт использования пакета PSPP для формирования профессиональных компетенции в области статистического анализа данных на основе доказательной медицины у обучающихся по направлению подготовки 31.05.01 «Лечебное дело» в ЮУГМУ. Практика показала эффективность использования пакета PSPP в медицинском вузе. К перспективам использования пакета PSPP в высшей школе следует отнести его функциональные возможности, достаточные для базового статистического анализа медицинской информации, и свободное использование.

ЛИТЕРАТУРА

1. Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 31.05.01 Лечебное дело (уровень специалитета) [Электронный ресурс]. — Режим доступа: <http://fgosvo.ru/news/8/1807> (дата обращения: 22.04.2019).
2. Темы практических занятий по дисциплине «Медицинская информатика» для обучающихся второго курса по специальности 31.05.01 Лечебное дело на 2018/2019 учебный год [Электронный ресурс]. — Режим доступа: http://www.chelsma.ru/files/misc/plan_lech_med.informatika_69468.pdf (дата обращения: 22.04.2019).
3. Основы доказательной медицины: учебное пособие / под общей ред. проф. Р. Г. Оганова. — М.: Силиция-Полиграф, 2010. — 136 с.
4. Флетчер Р., Флетчер С., Вагнер Э. Клиническая эпидемиология. Основы доказательной медицины. — М.: МедиаСфера, 1998. — 352 с.
5. Гринхальх Т. Основы доказательной медицины. — 4-е изд., перераб. и доп. — М.: ГЭОТАР-Медиа, 2015. — 336 с.
6. Реброва О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. — М.: Медиа Сфера, 2002. — 312 с.
7. Трухачёва Н. В. Математическая статистика в медико-биологических исследованиях с применением пакета Statistica. — М.: ГЭОТАР-Медиа, 2012. — 384 с.
8. Петри А., Сэбин К. Наглядная медицинская статистика: учеб. пособие / пер. с англ. под ред. В. П. Леонова. — 3-е изд., перераб. и доп. — М.: ГЭОТАР-Медиа, 2015. — 216 с.
9. PSPP — GNU Project — Free Software Foundation — Gnu.org [Электронный ресурс]. — Режим доступа: <https://www.gnu.org/software/pspp> (дата обращения: 22.04.2019).
10. Halter C. P. The PSPP Guide: An Introduction to Statistical Analysis. — 2th ed. — San Diego: CreativeMinds Press Group, 2017. — 194 p.
11. Bigham E. Introduction to PSPP: A Step by Step Guide. — Createspace Independent Pub., 2013. — 50 p.
12. Шамаева Т. Н., Степанова О. А. Научно-методическое обеспечение дисциплины «Информатика и медицинская статистика», реализующей программу аспирантуры в медицинском вузе // Вестник Томского государственного педагогического университета. — 2017. — № 12 (189). — С. 63–67.
13. Брико Н. И., Покровский В. И. Эпидемиологические исследования, клиническая эпидемиология и доказательная медицина // Медицинский альманах. — 2008. — № 5. — С. 15–19.
14. Фокин В. А., Карась С. И., Калипянская Т. А. Доказательная медицина в профессиональной подготовке врачей: информационные и экономические аспекты // Бюллетень сибирской медицины. — 2002. — № 4. — С. 47–61.

© Касюк Сергей Тимурович (sergey.kasyuk@gmail.com), Шамаева Татьяна Николаевна (shamtan@rambler.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»