

УДК 004:61

**ПРИМЕНЕНИЕ СТАТИСТИЧЕСКОГО ПАКЕТА PSPP
(PROGRAM FOR STATISTICAL ANALYSIS OF SAMPLED DATA)
В КУРСЕ «МЕДИЦИНСКАЯ ИНФОРМАТИКА»**

Степанова О.А., Диденко Г.А., Касюк С.Т.

*ФГБОУ ВО «Южно-Уральский государственный медицинский университет», Челябинск,
e-mail: okalst@mail.ru, pga80@mail.ru, sergey.kasyk@gmail.com*

Актуальность исследования обусловлена высокими требованиями ФГОС ВО для подготовки обучающихся медицинских вузов к применению статистических методов при решении профессиональных задач. Для будущих врачей важно владеть умениями применять методы математической статистики для обработки и интерпретации полученных результатов научно-исследовательской деятельности, для оценки и анализа состояния здоровья населения, анализа и оформления медико-статистической документации и отчетов в будущей профессиональной деятельности. Применение статистических пакетов позволит избежать ошибок в вычислениях. Работа исследователя фактически сводится к правильному выбору статистических методов и корректной интерпретации полученных результатов обработки. В статье обозначены функциональные возможности статистического пакета PSPP и представлена разработанная методика обучения методам математической статистики с применением данной программы в курсе медицинской информатики. Результаты исследования могут быть использованы для разработки концепции обучения медицинской информатике в вузе и, как следствие, совершенствования рабочей программы по этой дисциплине. Материалы статьи будут полезны преподавателям медицинской информатики. Цель статьи – раскрыть опыт обучения студентов медицинского вуза основным методам статистического анализа медицинских данных средствами статистического пакета PSPP.

Ключевые слова: дидактические возможности, статистический пакет PSPP, медицинская информатика, обучающиеся медицинского вуза

**APPLICATION OF THE STATISTICAL PACKAGE PSPP
(PROGRAM FOR STATISTICAL ANALYSIS OF SAMPLED DATA)
FOR COURSE «MEDICAL COMPUTER SCIENCE»**

Stepanova O.A., Didenko G.A., Kasyuk S.T.

*South Ural State Medical University, Chelyabinsk, e-mail: okalst@mail.ru,
pga80@mail.ru, sergey.kasyk@gmail.com*

The actuality of the research topic is caused by high requirements of the Federal standard of higher education for teaching medical students to use statistical methods for solving professional tasks. Professional skills of applying statistical methods will be important in future professional activities. These skills will allow: 1) processing and interpreting results of scientific researches; 2) evaluating and analyzing population health; 3) analyzing and processing medical and statistical documentations. Application of statistical packages will help to avoid errors in calculations. A task of a researcher is to choose statistical methods and interpret obtained results. This article analyses functionalities of the PSPP statistical package and gives the developed method of teaching mathematical statistics by using this package for the course «Medical computer science.» Results of the research can be used for developing concept learning of computer science in medical schools and perfecting syllabuses for the course. The paper can be of service to professors of medical schools. The goal of this article is to outline the experience of teaching medical students to apply basic methods of statistical analysis of medical data using the PSPP statistical package.

Keywords: didactic opportunities, PSPP statistical package, medical Informatics, medical school students

Основой доказательной медицины и современных клинических исследований являются статистические методы. Данные методы предлагают оптимальные модели для изучения медико-биологических процессов, позволяют описывать и критически оценивать эти процессы, выявлять закономерности и принимать обоснованные логические решения.

В статье представлен опыт обучения студентов медицинского вуза методам статистической обработки медицинских данных средствами статистического пакета PSPP (Program for Statistical Analysis of Sampled Data) с целью проведения оценки

и анализа состояния здоровья населения, обработки и представления результатов научно-исследовательской деятельности, оформления медико-статистической документации и отчетов в будущей профессиональной деятельности.

Материалы и методы исследования

Согласно ФГОС ВО выпускники медицинского вуза по специальностям 31.05.01 Лечебное дело и 31.05.02 Педиатрия должны освоить следующие компетенции [1, 2]:

– готовность к использованию основных физико-химических, математических и иных естественнонаучных понятий и ме-

тодов при решении профессиональных задач (ОПК-7);

– способность и готовность к применению социально-гигиенических методик сбора и медико-статистического анализа информации о показателях здоровья населения (ПК-4);

– готовность к анализу и публичному представлению медицинской информации на основе методов математической статистики и доказательной медицины (ПК-20);

– способностью к участию в проведении научных исследований (ПК-21).

В формировании указанных компетенций участвуют и такие дисциплины, как «Физика, математика» и «Медицинская информатика», которые относятся к базовым дисциплинам и обеспечивают преемственность в изучении статистических методов.

Отметим, что введение в учебный план дисциплины «Медицинская информатика» нового модуля «Стандартное и специализированное программное обеспечение для статистической обработки медико-биологических данных» инициировано внесенными изменениям в следующие нормативно-правовые акты: Федеральный закон от 29.07.2017 № 242-ФЗ «О внесении изменений в отдельные законодательные акты Российской Федерации по вопросам применения информационных технологий в сфере охраны здоровья» [3]; Приказ Минздрава России от 13.04.2018 № 165 (ред. от 28.09.2018) «Об утверждении плана информатизации Министерства здравоохранения Российской Федерации на 2018 год и на плановый период 2019 и 2020 годов» [4]; Профессиональный стандарт «Врач-лечебник (врач-терапевт участковый)», утвержденный приказом Министерства труда и социальной защиты РФ от 21.03.2017 г. № 293н [5]; Профессиональный стандарт «Врач-педиатр участковый», утвержденный приказом Министерства труда и социальной защиты РФ от 27.03.2017 № 306н [6].

На первом курсе, в рамках дисциплины «Физика, математика», обучающиеся изучают теоретические основы математической статистики, описательную статистику, основные статистические показатели и оценки выборки, получают представление о правилах формирования репрезентативной выборки. Ввиду ограниченного количества учебных часов, отведенных на дисциплину, не представляется возможным рассмотреть тему статистических методов в полном объеме. Однако сквозное изучение этой темы в рамках уже другой дисциплины, «Медицинская информатика», позволяет расширить знания по статистическим методам,

выполняя анализ данных в среде электронных таблиц и специализированном статистическом пакете.

Исходя из этого, значимым содержательным компонентом дисциплины «Медицинская информатика» является тема «Статистические методы анализа данных в медицине. Стандартное и специализированное программное обеспечение для статистической обработки медико-биологических данных». В процессе изучения данной темы в рамках лекционных и практических занятий рассматриваются следующие вопросы:

– Подготовка данных к статистическому анализу, группировка и кодирование данных.

– Типы статистических данных. Виды шкал, используемых для представления данных в медицине.

– Описательная статистика. Проверка данных на нормальность распределения.

– Статистическая проверка гипотез. Параметрические и непараметрические критерии. Сравнение зависимых и независимых выборок с помощью параметрических критериев.

– Однофакторный дисперсионный анализ. Непараметрический дисперсионный анализ.

– Исследование зависимостей: корреляционный анализ.

– Регрессионный анализ. Линейный регрессионный анализ. Бинарная логистическая регрессия. Множественная линейная регрессия.

– Многомерные статистические методы (на примере кластерного анализа).

На практических занятиях обучающиеся приобретают умения правильного представления медицинских данных, их группировки и ввода в программной среде, умения рационально выбирать и применять статистические методы, анализировать и интерпретировать полученные результаты. Для погружения в тему обучающиеся изучают возможности статистической обработки данных не только в среде электронных таблиц, которые имеют ограничения по выбору методов, но специализированных статистических пакетов на примере пакета PSPP (Program for Statistical Analysis of Sampled Data).

Отметим, что применение специализированных статистических пакетов возможно только при условии, что обучающиеся владеют основными приемами работы в программных средах, знают теоретические основы статистической обработки данных, назначение и границы применения основных статистических методов. Все это

позволит им правильно интерпретировать полученные результаты.

Статистические программы условно делят на профессиональные, универсальные и специализированные, также их можно классифицировать как бесплатно распространяемые и коммерческие. К свободному программному обеспечению (Free Software) относятся достаточно большое количество универсальных статистических пакетов: PSPP, Macanova, Vista, MyStat, PAST, язык R.

В нашем курсе была выбрана кроссплатформенная универсальная статистическая программа с закрытым кодом PSPP, разработанная в рамках проекта GNU (General Public License) [7].

Выбор данного пакета был обусловлен прежде всего тем, что он является альтернативой статистической программы для социальных исследований IBM SPSS Statistics. Указанные программы имеют идентичный экранный интерфейс и принципы организации работы, а также совместимость данных. Поскольку программа PSPP находится в свободном доступе, то это позволит обучающимся самостоятельно обрабатывать результаты научно-исследовательской работы на своих личных компьютерах. В дальнейшем, для расширенного анализа данных, обучаясь в ординатуре и аспирантуре, обучающиеся смогут без особых трудностей перейти и использовать уже профессиональный пакет SPSS, который выполняет всевозможные сложные виды статистического анализа в медицине.

Результаты исследования и их обсуждение

На первом практическом занятии обучающиеся знакомятся с интерфейсом программы PSPP, осваивают процедуру описания переменных. Экранный интерфейс программы PSPP, как и SPSS, представлен окном просмотра переменных, окном просмотра данных и окном вывода результатов.

Статистический анализ включает несколько этапов, и одним из важных является правильное описание переменных, адекватное кодирование и группировка данных.

Так, достаточно часто в медицинских исследованиях возникает необходимость перевести качественные данные в количественные. Например, пол для удобства обработки представить в дихотомической шкале, а возраст отнести к определенной возрастной группе и представить в порядковой шкале. Программа PSPP позволяет редактировать уже описанные и введенные

переменные: перекодировать переменные в новые, выполнять фильтрацию данных.

Проверка данных на соответствие нормальному закону распределения является важным этапом статистического анализа, так как позволяет в последующем корректно выбирать статистический критерий. Проверка на нормальность распределения осуществляется с помощью вычисления описательных характеристик, представления данных в графическом виде, а также применения статистических критериев (Колмогорова – Смирнова, Хи-квадрат).

После выполнения команды по статистической обработке данных в программе PSPP открывается окно просмотра результатов, в котором представлены результаты в табличном и графическом виде. Основные трудности у обучающихся возникают именно на этом этапе, когда необходимо правильно интерпретировать полученное решение и сформулировать корректные выводы. Поэтому обучающимся в задании предлагается подробное и полное описание таблиц, графиков и необходимых статистических правил принятия и отклонения гипотез.

Предмет математической статистики является достаточно сложным для обучающихся медицинского вуза, так как большинство из них сдает только базовый уровень ЕГЭ по математике. Для того чтобы методы статистики не воспринимались абстрактно, для активизации познавательного интереса к предмету и восприятию статистики, как важного инструмента доказательной медицины и проведения клинических испытаний, практические задания по теме подобраны с усилением профессионально-прикладной направленности.

В ходе выполнения заданий делается акцент не на конкретные статистические вычисления, а на понимание обучающимися возможностей и логики того или иного статистического метода, требований и ограничений, предъявляемых к его применению, на анализ полученного результата.

Рассмотрим дидактическую логику построения заданий на примере темы «Регрессионный анализ».

Задание. Определить параметры линейного регрессионного уравнения:

$$choll = b \cdot chol0 + a,$$

где *choll* – показатель холестерина через один месяц после начала приема препарата пациентами; *chol0* – начальный уровень холестерина. После определения параметров *b* и *a*, зная исходный показатель холестерина, можно спрогнозировать значение холестерина, которое будет через один месяц.

1. Описать переменные и ввести данные.
2. Выбрать в меню Анализ – Регрессия – Линейная.

3. Задать переменную *chol1* (Холестерин через 1 месяц) в поле для Зависимые и переменную *chol0* (Холестерин начальное) в Независимые.

4. Выбрать параметры: коэффициенты, доверительный интервал, *R*.

5. Сохранить и выбрать параметры: прогнозные значения и остатки.

В окне вывода PSPP появятся две таблицы.

Таблица «Резюме модели» содержит следующие показатели (табл. 1): коэффициент корреляции $R = 0,85$, который свидетельствует о тесной и прямой связи между переменными *chol0* и *chol1*; коэффициент детерминации $R^2 = 0,72$ показывает долю общей вариации, которую способна объяснить регрессионная модель. Доля вариации зависимой переменной является одним из показателей качества модели, чем ближе R^2 к 1, тем качественнее построенная модель.

Статистическая значимость модели оценивается с помощью критерия Фишера (*F*): если достигнутый уровень значимости критерия (0,00) меньше 0,05, то принимается альтернативная гипотеза H_1 , т.е. подтверждается статистическая значимость модели.

Следующая таблица «Коэффициенты» содержит показатели, характеризующие коэффициенты регрессионного уравнения (табл. 2).

Получены следующие коэффициенты уравнения регрессии: $a = 71,93$; $b = 0,71$.

Коэффициенты являются статистически значимыми по *t*-критерию, поскольку достигнутые уровни значимости меньше 0,05 [8, 9]. Таким образом, можно записать уравнение регрессии:

$$chol1 = 0,71 \cdot chol0 + 71,93.$$

В следующем задании студентам предлагается самостоятельно выполнить аналогичный анализ зависимости между начальным уровнем холестерина и уровнем холестерина через шесть месяцев после начала приема препарата пациентами, а также проанализировать и сделать выводы. Для наглядности результаты оформляются в виде таблицы, что дает возможность обучающимся структурировать и систематизировать свои знания.

Задание. Переменные *chol0*, *chol6* представляют начальный уровень холестерина и уровень холестерина через 6 месяцев после начала приема определенного препарата пациентами. Выяснить, как коррелируют между собой эти показатели. Определить параметры линейного регрессионного уравнения, принимая переменную *chol6* как зависимую переменную (*y*), а переменную *chol0* – как независимую переменную (*x*).

1. Для переменных *chol0*, *chol6* построить гистограммы, включающие кривые нормального распределения.

2. Сделать вывод о соответствии данных *chol0*, *chol6* нормальному закону распределения и оформить отчет в табл. 3.

Таблица 1

Линейная регрессия. Резюме модели (Холестерин, через 1 м)

R	R квадрат	Скорректированный R квадрат	Ст. погрешность оценки
0,85	0,72	0,71	23,16

Таблица 2

Коэффициенты регрессионного уравнения. Коэффициенты (Холестерин, через 1 месяц)

	Нестандартизованы коэффициенты		Стандартизованные коэффициенты	t	Знач.	95%-ный доверительный интервал для В	
	В	Станд. погрешность	Бетта			Нижняя граница	Верхняя граница
(Константа)	71,93	20,50	0,00	3,51	0,001	29,95	113,92
Холестерин, начальное	0,71	0,08	0,85	8,57	0,000	0,54	0,88

Таблица 3

Отчет по критерию Колмогорова – Смирнова

	chol0	chol6
Значение критерия Колмогорова – Смирнова		
Достигнутая значимость критерия		
Вывод о соответствии изучаемого распределения нормальному распределению		

Таблица 4

Отчет по F-критерию и коэффициентам регрессионного уравнения

	<i>chol0 – chol6</i>	Значимость
Значение критерия Фишера <i>F</i>		
Вывод о статистической значимости модели		
Значение коэффициента корреляции, <i>R</i>		
Вывод о виде связи		
Значение коэффициента детерминации, <i>R</i> ²		
Вывод о доле вариации и качестве модели		
Значение <i>a</i>		
Значение <i>b</i>		
Вывод о статистической значимости константы и коэффициента регрессии		
Полученное уравнение регрессии		
Значение холестерина <i>chol6</i> , если <i>chol0</i> = 199		

Таблица 5

Отчет по критерию Колмогорова – Смирнова

Значение критерия Колмогорова – Смирнова	
Достигнутая значимость критерия	
Вывод о нормальном распределении остатков	

3. Определить параметры линейного регрессионного уравнения, рассматривая показатель холестерина через 6 месяцев (переменная *chol6*) как зависимую переменную, а начальное значение холестерина (переменная *chol0*) – как независимую переменную. Сделать вывод и оформить отчет в табл. 4.

4. Проверить распределение остатков модели на соответствие нормальному закону распределения. Сделать вывод и оформить отчет в табл. 5.

Заключение

Таким образом, для того чтобы реализовать образовательный потенциал статистических пакетов при изучении методологии статистики, была использована специальным образом разработанная система практических заданий, которые выстроены в строгой последовательности: от репродуктивных заданий с подробным описанием алгоритма работы, с акцентированием на интерпретацию полученных результатов, до полностью самостоятельных заданий [10, с. 81]. При разработке заданий особое внимание уделялось их профессиональной направленности. Применение статистических пакетов в обучении, выстроенное на принципах профессиональной направленности и междисциплинарной интеграции, способствует повышению мотивации обучающихся, пониманию необходимости знаний методов математической статистики для будущей профессиональной

деятельности и в целом формирует системное представление о статистическом моделировании в медицине.

Безусловно, как свободное программное обеспечение, пакет PSPP имеет ряд ограничений в своем функционале: отсутствие экспорта и импорта данных, ограниченный набор статистических критериев. Однако даже при этих недостатках статистический пакет PSPP позволяет сформировать у обучающихся целостное представление об этапах проведения статистического исследования, о подготовке данных к обработке, о процедуре выполнения основных статистических методов. И самое главное, позволяет обучающимся приобрести умения интерпретации полученных результатов обработки данных.

В заключение отметим, что пролонгированное изучение методов математической статистики с поддержкой статистических пакетов позволяет оптимизировать процесс изучения статистики в медицинском вузе.

Список литературы

1. Утвержден ФГОС ВО по направлению подготовки 31.05.01 Лечебное дело (далее соответственно – программа специалитета, специальность). [Электронный ресурс]. URL: <http://fgosvo.ru/news/2/1807> (дата обращения: 12.05.2021).
2. Утвержден ФГОС ВО по направлению подготовки 31.05.02 Педиатрия (далее соответственно – программа специалитета, специальность). [Электронный ресурс]. URL: <http://fgosvo.ru/news/21/1327> (дата обращения: 13.05.2021).
3. Приказ Минздрава России от 13.04.2018 № 165 (ред. от 28.09.2018) «Об утверждении плана информатизации Министерства здравоохранения Российской Федерации»

на 2018 год и на плановый период 2019 и 2020 годов». [Электронный ресурс]. URL: http://www.consultant.ru/document/cons_doc_LAW_298642/ (дата обращения: 13.05.2021).

4. Федеральный закон «О внесении изменений в отдельные законодательные акты Российской Федерации по вопросам применения информационных технологий в сфере охраны здоровья» от 29.07.2017 № 242-ФЗ (последняя редакция). [Электронный ресурс]. URL: http://www.consultant.ru/document/cons_doc_LAW_221184/ (дата обращения: 14.05.2021).

5. Об утверждении профессионального стандарта «Врач-лечебник (врач терапевт-участковый)». [Электронный ресурс]. URL: <http://fgosvo.ru/uploadfiles/profstandart/02.009.pdf> (дата обращения: 14.05.2021).

6. Об утверждении профессионального стандарта «Врач-педиатр участковый». [Электронный ресурс]. URL: <http://fgosvo.ru/uploadfiles/profstandart/02.008.pdf> (дата обращения: 15.05.2021).

7. Obtaining GNU PSPP. [Электронный ресурс]. URL: <http://www.gnu.org/software/pspp/get.html> (дата обращения: 15.05.2021).

8. Гржибовский А.М., Иванов С.В., Горбатова М.А. Однофакторный линейный регрессионный анализ с использованием программного обеспечения Statistica и SPSS // Наука и Здоровоохранение. 2017. № 2. С. 5–33.

9. Шарашова Е.Е., Холматова К.К., Горбатова М.А., Гржибовский А.М. Применение множественного линейного регрессионного анализа в здравоохранении с использованием пакета статистических программ SPSS // Наука и Здоровоохранение. 2017. № 3. С. 5–31.

10. Степанова О.А., Шамаева Т.Н. Научно-методическое обеспечение дисциплины «Информатика и медицинская статистика», реализующей программу аспирантуры в медицинском вузе // Вестник Томского государственного педагогического университета. 2016. № 8 (173). С. 79–82.