

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ В СИСТЕМЕ ВЫСШЕГО ПСИХОЛОГИЧЕСКОГО ОБРАЗОВАНИЯ

УДК 159.9.07

Светлана Викторовна Смирнова,
к.п.с.н., доц. каф. психологии и педагогики,
Амурский государственный университет
(АмГУ), г. Благовещенск
Тел.: (914) 564-66-44
Эл. почта: smirnova2001@mail.ru

Татьяна Анатольевна Макаrchук
к.п.н., доц. каф. Информатики, Санкт-
Петербургский государственный универ-
ситет экономики и финансов (СПбГУЭФ),
г. Санкт-Петербург
Тел.: (931) 207-92-56
Эл. почта: tmakarchuk@mail.ru

В статье рассматриваются особенности применения информационных технологий в рамках использования прикладной статистики в психологии. Анализируются требования к статистической подготовке студентов-психологов в условиях информационного общества.

Ключевые слова: информационные технологии анализа данных, прикладная статистика, психометрика, статистическое мышление.

Svetlana V. Smirnova,
PhD in Psychological Sciences, Associate
Professor, the Department of Psychology
and Pedagogics, Amur State University,
Blagoveshchensk city
Tel.: (914) 564-66-44
E-mail: smirnova2001@mail.ru

Tatyana A. Makarchuk,
PhD in Pedagogic Sciences, Associate
Professor, the Department of Informatics,
St. Petersburg University of Economic and
Finance (FINEC), Saint-Petersburg
Tel.: (931) 207-92-56
E-mail: tmakarchuk@mail.ru

INFORMATION TECHNOLOGIES OF THE STATISTICAL DATA ANALYSIS WITHIN THE SYSTEM OF HIGHER PSYCHOLOGICAL EDUCATION

The features of using information technologies within applied statisticians in psychology are considered in the article. Requirements to statistical preparation of psychology students in the conditions of information society are analyzed.

Keywords: information technologies of data analysis, applied statistics, psychometrics, statistical thinking.

1. Введение

Мошное развитие информационных технологий анализа данных способствует расширению сферы прикладных статистических исследований в медицине и психологии. Прикладная статистика становится необходимым исследовательским инструментом психолога для обеспечения его конкурентоспособности. «Статистическое мышление» и владение статистическими методами исследования обязательны для принятия оптимальных стратегических решений в психокоррекции и психотерапии.

Требования ФГОС ВПО специальности 030401.65-клиническая психология и направлений 030300.62- психология и 050400.62- психолого-педагогическое образование к уровню подготовки будущих психологов в области информационных технологий анализа данных следует рассматривать в нескольких аспектах.

Научно-технические требования к подготовке бакалавров включают: моделирование, прогнозирование, оценку надежности и качества исследования. Научно-технические требования нашли свое отражение в системе ГОСТ Р ИСО 9001-2001, основанных на статистических методах управления для оценки качества, сертификации и классификации продукции и услуг [1].

Социально-экономические требования к психологам в области информационных технологий анализа данных включают освоение правил и норм базовых социальных практик в организации прикладных исследований, экономическую оценку их эффективности.

Духовно-нравственные требования характеризуются становлением статистического мышления и развитием информационной культуры.

2. Структура статистического анализа данных в психологии

На рисунке 1 представлена иерархическая структура статистических модулей, в разной степени реализованных в компьютерных пакетах анализа данных.

Методологический фундамент прикладной статистики в психологических исследованиях составляет система научных категорий, теоретических положений и методов статистического исследования в направлениях: представление данных, анализ данных, моделирование.

Развитие информационных технологий анализа данных, отражающие содержание прикладной статистики в психологии, происходит в направлениях: математико-статистического описания числовых (случайных) величин; математического моделирования случайных процессов; статистического анализа и обработки объектов нечисловой природы.

3. Возможности использования современных пакетов программ в психологии

На российском рынке программных продуктов анализа статистических данных присутствует как иностранные, так и отечественные поставщики. Обилие пакетов статистических программ, которых по данным Международного статистического института (<http://isi-web.org>) насчитывается более 1000, приводит к усложнению процесса выбора оптимальной программы для отдельного исследования.

Западные программные продукты анализа данных, относящиеся к наукоемкому программному обеспечению, образуют многофункциональные пакеты высшего ценового класса (\$ 1-4 тыс. на одну рабочую станцию в бизнес-структуре, \$ 200-600 на один ПК для студентов и исследователей, работающих со средними размерами выборки). Примерами таких программ, наиболее используемыми в российском секторе экономике, являются SAS (www.sas.com), STATA (www.stata.com), IBM SPSS STATISTICS (www.spss.com), Statistica (www.statsoft.com), Mplus (<http://www.statmodel.com>) и др.

Отечественные разработки статистических пакетов появились в 80-90-е годы XX века и преимущественно были представлены ведущими вузами страны (STADIA, ОЛИМП: Статэксперт и др.). Российские производители предлагают



Рис. 1. Иерархическая структура реализации статистических модулей в компьютерных пакетах анализа данных

свои программные продукты анализа данных, превосходящие западные по двум основным параметрам – доступным ценам (профессиональная версия до \$300) и простоты интерфейса. Среди статистических пакетов российских разработок в настоящее время хорошо известна STADIA (<http://statsoft.msu.ru>).

К программным продуктам, способным обрабатывать большие массивы данных, но обладающих узким спектром статистических процедур и слабой графикой, относятся математические пакеты MatLab, Mathematica, MathCad, Maple и др., модуль «Анализ данных» в Excel. Авторитетные психологические журналы часто не признают статистические расчеты, выполненные в данных программах.

Высокая стоимость профессиональных и полупрофессиональных пакетов программ анализа данных формирует спрос на свободное программное обеспечение в области анализа данных, пригодное для студенческих научных исследований с небольшими объемами выборки. На сайте www.freestatistics.info представлено программное обеспечение типа open source software – открытое ПО, в ко-

торых права пользователя («свободы») на неограниченные установку, запуск, а также свободное использование, изучение, распространение и изменение программ защищены юридически авторскими правами при помощи свободных лицензий. Примерами таких программ являются: DATAPLOT, INSTAT+, R, StatistX и др. Массовость использования пакета статистических программ формируется с учетом пол-

ноты и качества реализованных в пакете статистических методов, наличия возможностей экспорта/импорта данных, преобразования данных, графического анализа данных, встроенного языка программирования и др.

Например, информационные технологии графического анализа и представления данных наиболее удачно реализованы в программе Statistica. На рис. 2 представлен двумерный

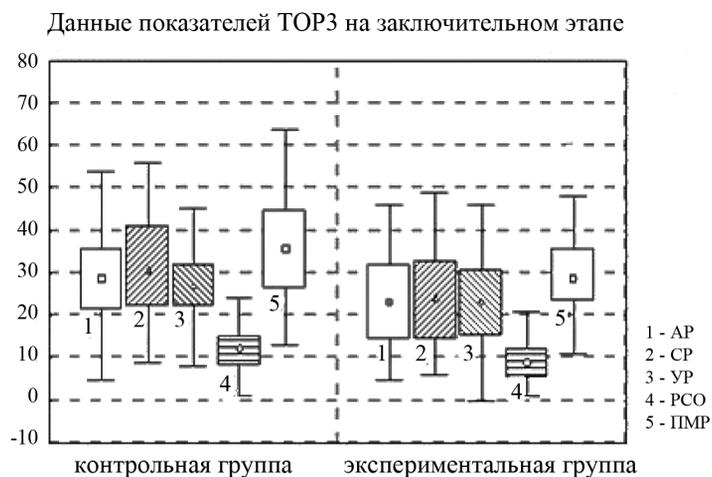


Рис. 2. 2D-Box plot в программе Statistica

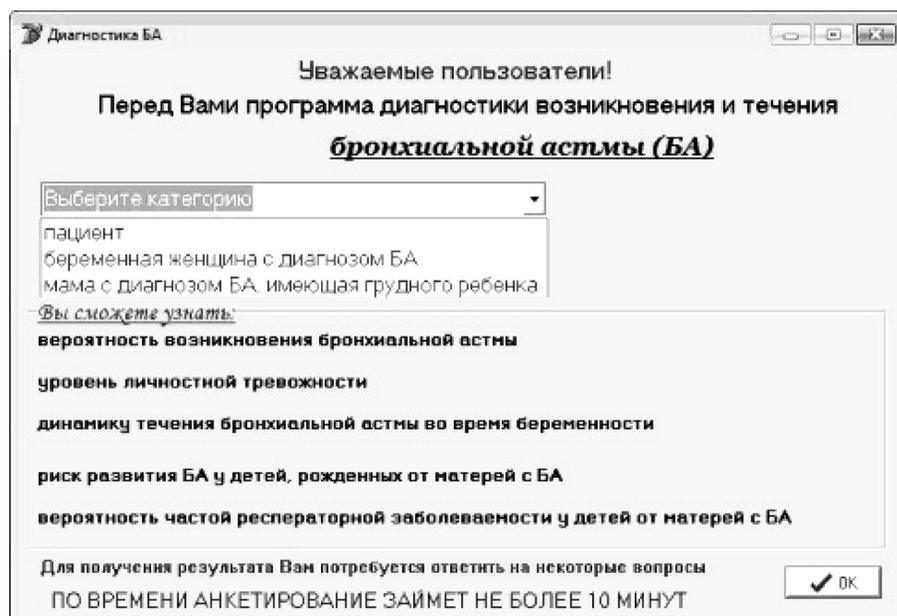


Рис. 3. Фрагмент стартового окна программы «Диагностика развития бронхиальной астмы у детей»

группирующий график типа 2D-Box Plot, позволяющий наглядно отобразить результаты непараметрической статистики, в частности критерия U - Манна-Уитни выявления различий в двух группах (результаты получены в рамках лонгитюдного экспериментального исследования студенческих выборок по Томскому опроснику ригидности Г.В. Залевского)[2].

4. Использование методов статистического исследования в психометрике

Психометрика – это научная дисциплина, которая дает количественное выражение взаимосвязей, взаимозависимостей психологических явлений и процессов. Психометрика с помощью методов статистики позволяет компактно описать данные, информацию, понять их структуру, провести классификацию и выявить закономерности.

Объектом изучения психометрики являются количественные данные таких типов как перекрестные (пространственные) данные, временные ряды и панельные данные, для анализа которых используются следующие модели: модель множественной регрессии; временные ряды; системы регрессионных уравнений; моделирование панельных данных. Классификация типов и моделей данных позволяет выделить содержательную направленность развития информационных

технологий анализа психологических данных.

С помощью применения статистических программ в психологии и клинической практики, возможно создание моделей прогнозирования вероятности развития тех или иных расстройств. Примером такого прогнозирования является модель, построенная с использованием совокупности инструментальных средств диагностики течения бронхиальной астмы у детей [3].

Данные исследования были оформлены в виде компьютерной программы, предлагающей пройти несложное тестирование в условиях поликлиники (рис. 3).

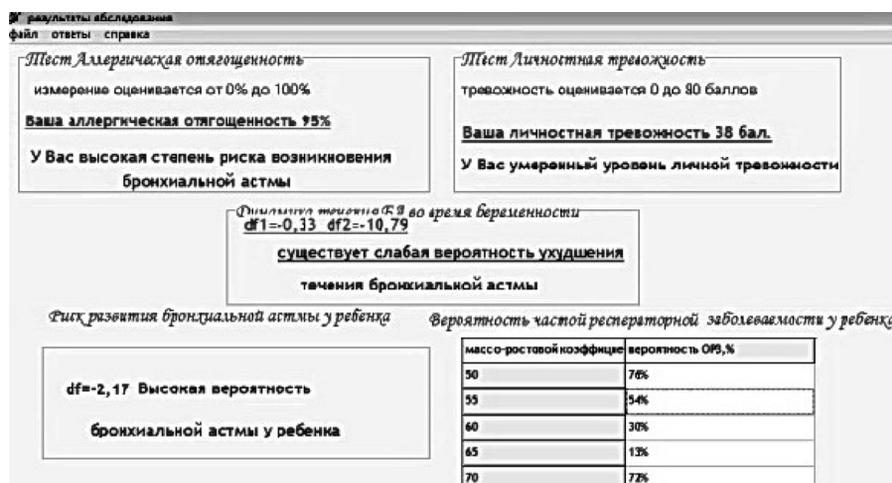


Рис. 4. Фрагмент окна программы результатов моделирования развития бронхиальной астмы

В ходе исследования были использованы:

- Тест тревожности Спилбергра-Ханина;
- Данные медицинских исследований;
- Многомерные методы исследования, в т.ч. корреляционный, дискриманантный, регрессионного пробит-анализ;
- Информационные технологии статистического анализа данных;
- Технологии программирования.

Результаты выполнения программы ««Диагностика развития бронхиальной астмы у детей» представлены на рис. 4.

Психолог, владеющий информационными технологиями статистического анализа данных способен ориентироваться в множестве абстрактных показателей для оценки объективного психического и психологического статуса обследуемого, или групп респондентов как в рамках номотетической, так и в рамках идеографической парадигмы [4].

Применение информационных технологий в рамках психологического исследования студентами в процессе обучения в вузе характеризуется, с одной стороны, кажущейся простотой их анализа, с другой стороны, полученные результаты часто сопровождаются большим количеством ложных выводов.

Первая особенность применения информационных технологий анализа данных связана с проблемой сбора данных и их достоверности.

Вторая особенность анализа психологических данных связана с возникновением ложной корреляции при

Таблица 1.

Корреляционная матрица показателей, выполненная в Statistica 5.0.

	<i>V</i>	<i>I</i>	<i>H</i>	<i>B</i>
<i>V</i>	1,00	0,99	0,99	-0,99
<i>I</i>	0,99	1,00	0,99	-0,99
<i>H</i>	0,99	0,99	1,00	-0,98
<i>B</i>	-0,99	-0,99	-0,98	1,00

изучении взаимосвязей между психологическими переменными. Ошибочное восприятие тесной зависимости между переменными, не имеющими не только прямой, но и нелинейной зависимости между собой, объясняется наличием общих причин, влияющих на обе переменные.

В таблице 1 представлена корреляционная матрица показателей развития Дальневосточного региона в период с 2004-2011 гг., не имеющие явной прямой зависимости между собой, таких как:

V – внутренние затраты на исследования и разработки, млн. руб.;

I – инвестиции в основной капитал, млн. руб.;

H – средние цены на первичном рынке жилья, тыс. руб.;

B – число больничных коек, тысяч.

Ложность выводов, получаемых при анализе таблице 1, типа тесной обратной зависимости между ценами на первичном рынке жилья и числом больничных коек ($r_{H,B} = -0,98$), нашла психологическое обоснование и в дальнейшем сопровождалась ошибочными прогнозами при построении регрессии.

Выявить ложную корреляцию можно при помощи оценки корреляционной связи между первыми разностями этих показателей, при этом информационные технологии позволяют получить новый временной ряд с лагом один (два и т.д.) за достаточно короткое время. Переход к лаговым переменным позволяет в ряде случаев избежать мультиколлинеарности при

построении модели множественной регрессии [5].

Среди других особенностей использования информационных технологий анализа клинко-психологических данных можно выделить: массовое использование параметрической статистики, в частности коэффициента корреляции Пирсона, без проверки наличия нормального распределения случайной величины; сложности построение нечеткой множественной регрессии для анализа интервальных данных; выбор случайных и фиксированных эффектов при построении модели панельных данных и др.

5. Заключение

Информационные технологии анализа данных, ставшие обязательным инструментом плановых, аналитических, маркетинговых отделов бизнес-структур и правительственных учреждений, постоянно совершенствуются, становясь доступными большому количеству пользователей, при этом требования к их «статистической» компетентности возрастают. Для установления статистических закономерностей и определения эффективности проведенного воздействия, психолог должен имеет возможность динамической оценки показателей как в психосоциальном, так и клинко-психологическом контексте.

Литература

1. Система международных стандартов ISO серии 9000 версии 2000 года. Электронный ресурс: http://quality.eup.ru/GO_ST/st.htm <<http://quality.eup.ru/GOST/st.htm>>

2. Смирнова С.В., Залевский Г.В. Профилактика дезадаптации первокурсников вуза посредством развития их психической гибкости // Сибирский психологический журнал. 2005. № 22. С. 54-58.

3. Арутюнян К.А., Бабцева А.Ф., Макаrchук Т.А. Способ прогнозирования рецидивирующего течения обструктивного бронхита у детей раннего возраста с перинатальной энцефалопатией. Патент на изобретение RUS 2430677 23.03.2010

4. Смирнова С.В. Супервизия в процессе профессионального становления личности специалистов-психологов// Сибирский психологический журнал. 2008. № 30. С. 49-50.

5. Макаrchук Т.А., Юрьева Т.А., Лебедь О.А. Методологические проблемы обучения прикладной статистике бакалавров гуманитарных направлений // Гуманизация образования. 2009. № 5. С. 28-33.

References

1. The system of the international standards ISO of a series 9000 version 2000. The electronic resource: http://quality.eup.ru/GO_ST/st.htm <<http://quality.eup.ru/GOST/st.htm>>

2. Profilaktika of disadaptation of first-year students of university at the expense of development of their mental flexibility//The Siberian psychological magazine. 2005 . No. 22. Page 54-58.

3. Harutyunyan K.A., Babtseva A.F., Makarchuk TA

3. Arutyunyan K.A. Babtseva A.F. Makarchuk T.A. The method of prediction of obstructive bronchitis in children of early age with perinatal encephalopathy. The patent for the invention RUS 2430677 23.03.2010

4. Smirnova S. V. Supervizy in the course of professional formation of the identity of specialists psychologists//the Siberian psychological magazine. 2008 . No. 30. Page 49-50.

5. Makarchuk TA, Yuryeva T., Lebed O.A. Methodological problems of training in applied statistics of bachelors of the humanitarian directions// education Humanization. 2009. No. 5. Page 28-33.