

ханізмів формування капіталу із випуску цінних паперів, та маючи майнову відокремленість від КУА, яка виконує функції управління щодо них.

Бова Андрей Андреевич

*кандидат социологических наук, старший научный сотрудник,
начальник научно-исследовательского отдела Государственного
научно-исследовательского института МВД Украины
ORCID: 0000-0003-1588-0250*

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ПРОВЕДЕНИЯ И АНАЛИЗА РЕЗУЛЬТАТОВ СОЦИОЛОГИЧЕСКИХ ОПРОСОВ

В докладе рассматривается использование программного обеспечения на разных этапах социологического исследования. Для анализа и визуализации данных применяются Microsoft Excel с надстройками, универсальные статистические пакеты (IBM SPSS Statistics, Stata, SAS, TIBCO Statistica), специализированные программы (IBM SPSS Amos, Mplus) и платформы бизнес-аналитики (Microsoft Power BI, Tableau). Приводятся тренды использования программного обеспечения в социальных исследованиях.

Ключевые слова: социологический опрос, программное обеспечение, выборка, сбор данных, статистический анализ, визуализация данных.

Bova Andriy

*Ph.D., Senior Researcher, Head of the Research Department of the State
Research Institute of the Ministry of Internal Affairs of Ukraine
ORCID: 0000-0003-1588-0250*

SOFTWARE FOR CONDUCTING AND ANALYZING OF RESULTS OF PUBLIC OPINION SURVEY

The report discusses the use of software at different stages of a case study. For data analysis and visualization, Microsoft Excel with add-ins, general-

purpose statistical packages (IBM SPSS Statistics, Stata, SAS, TIBCO Statistica), specialized purpose programs (IBM SPSS Amos, Mplus) and business intelligence platforms (Microsoft Power BI, Tableau) are used. Trends in the use of software in social research are presented.

Keywords: opinion poll, software, sample size, data collection, statistical analysis, data visualization.

Социологические исследования, основывающиеся на количественных методах, направлены либо на подтверждение теоретических положений (иногда конкурирующих между собой), либо на выявление эмпирических закономерностей и дальнейшую их содержательную интерпретацию. Методология и методика социологического исследования включают в себя определение цели и задач, объекта и предмета, формулировку гипотез, выбор методов сбора информации, формирования выборочной совокупности, избрание методов обработки данных и применения соответствующих программных продуктов.

Программное обеспечение задействуется на таких этапах социологического исследования: расчет выборки и мощности статистического критерия, сбор и ввод данных, предварительная обработка, анализ данных и их визуализация.

Специализированные программы применяются для определения необходимого объема выборки респондентов/испытуемых (калькуляторы выборки), в том числе с учетом мощности критерия (G*Power, модуль SPSS Complex Samples программы IBM SPSS Statistics, PASS, язык программирования для статистической обработки данных и работы с графикой R, SAS, TIBCO Statistica).

Сбор эмпирической информации может происходить путем интервьюирования «лицом к лицу» с использованием планшетов (для этого разработано программное обеспечение OCA CAPI Android, DigSee SURE) или заполнения бумажных бланков, раздаточного анкетирования, телефонных опросов (OCA CATI), интернет-опросов (сервисы Google Forms, LimeSurvey, EsurveysPro), рассылки электронных форм анкет на почту, посредством использования приложений, которые устанавливаются на гаджеты (мобильные телефоны, планшеты).

Ввод собранных в процессе изучения общественного мнения данных в компьютер осуществляется с помощью Microsoft Excel, модуля OCA Make Form программы OCA for Windows, программы EpiData Entry. Этап подготовки данных непосредственно к анализу включает проведение логического контроля, очистку данных, формирование новых признаков и интегральных индексов, для чего используются электронные таблицы, языки программирования, универсальные статистические пакеты.

Распространенность того или иного программного продукта обуславливается многими факторами: возможностями для решения определенных задач, исследовательскими потребностями, лицензионной политикой, корпоративными стандартами, уровнем технической поддержки, обновлением продукта, наличием сообщества пользователей, учебных курсов, научно-методической литературы и др. Отметим, что стоимость программного обеспечения примерно одного класса различных производителей может существенно колебаться. Она также зависит от типа лицензии (постоянная или временная), целей использования и статуса организации, модулей, входящих в комплект и т. п. Существуют различные подходы к определению популярности программного обеспечения: рейтинги консалтинговых агентств (квадранты Gartner) и ревью-площадок (G2 Crowd), опросы пользователей, в том числе, представителей научных кругов, подсчет количества ссылок в научных публикациях, требования к кандидатам на рынке труда и др.

Анализ данных включает построение таблиц и графиков, исследование числовых характеристик переменных: процентов, мер центральной тенденции (среднее, мода, медиана), мер положения (квантили), рассеяния (дисперсия, стандартное отклонение, коэффициент вариации) и формы (асимметрия, эксцесс) распределений. Поскольку большинство социологических опросов проводится на выборке респондентов, необходимо использовать выводную статистику, с помощью которой результаты анализа переносятся на генеральную совокупность.

Взаимосвязь между признаками устанавливается на основе мер связи, отличия распределений в группах – с помощью однофактор-

ного дисперсионного анализа и однофакторного рангового дисперсионного анализа. Многомерные методы включают в себя общие и обобщенные линейные модели, деревья регрессий и классификаций, логлинейный анализ, методы редукции признаков пространства (факторный анализ, множественный анализ соответствий), иерархические и неиерархические процедуры кластерного анализа, латентно-классовый анализ, многомерное шкалирование и пр.

В социологии большинство переменных измеряются в номинальных или порядковых шкалах. К последним в ряде случаев применяются параметрические методы статистики. Точность измерения социальных показателей улучшается при использовании интегральных индексов как взвешенной суммы баллов переменных, имеющих между собой хорошую внутреннюю согласованность.

Объяснение социальных явлений через конструкторы, оцениваемые с помощью факторного анализа, позволяет выдвинуть или проверить теоретические гипотезы о структуре взаимосвязей между несколькими переменными, обнаружить небольшое число латентных факторов, для которых применим метрический уровень измерения.

Для переменных, измеренных в порядковой шкале, которые включаются в различные процедуры снижения размерности, корреляционная матрица рассчитывается на соответствующих мерах связи – полихорической корреляции, коэффициенте ранговой корреляции Спирмена, коэффициенте ранговой корреляции Кендалла. Ординальные переменные также могут быть подвергнуты предварительной оцифровке или преобразованию с учетом байесовских оценок.

Метод частичных наименьших квадратов (partial least squares) используется для решения задач снижения размерности, регрессии и классификации, в моделировании структурными уравнениями. Его специфика заключается в построении зависимостей в тех случаях, когда предикторы сильно коррелируют между собой или когда размер выборки невелик. В кросс-культурных исследованиях получили распространение многоуровневые регрессионные модели, позволяющие исследовать закономерности в выборках принимая во внимание иерархический характер данных.

Для обработки и анализа эмпирической информации применяются различные программные продукты. В частности, анализ данных возможен с помощью офисного приложения Microsoft Excel, а также целого ряда надстроек к нему: «Анализ данных», AtteStat, Real Statistics, Sigma Magic, Statistician, StatPlus, XLSTAT. В Украине службы по изучению общественного мнения широко используют в своей деятельности пакеты OCA for Windows и OCA New Line.

Анализ массивов социологической информации, как правило, осуществляется с помощью универсальных статистических пакетов. По результатам изучения публикаций в поисковой платформе Google Академия за 2018 г. установлено, что при написании научных работ на сегодняшний день чаще всего используются SPSS Statistics, SAS, R, Stata [1, Figure 2a].

В течение двадцати лет лидерство остается за программой SPSS, принадлежащей в настоящее время компании IBM [1, Figure 2d]. Неоспоримыми преимуществами данного пакета является дружелюбный графический интерфейс пользователя, возможности по управлению и статистическому анализу данных, наличие командного языка (синтаксиса).

SAS и Stata обладают большими возможностями по сравнению с SPSS в программировании пользовательских приложений.

Для моделирования структурными уравнениями и, в частности, подтверждающего факторного анализа используются специализированные пакеты IBM SPSS Amos, LISREL, Mplus, SmartPLS, а для латентно-классового анализа – IBM SPSS Amos, Latent GOLD и Mplus.

Следует отметить, что кроме коммерческих программных продуктов, существует также и свободно распространяемое программное обеспечение для решения задач обработки данных социологических опросов, например, R, SAS University Edition, PSPP.

Количество ссылок в научных публикациях, индексируемых Google Академия, на IBM SPSS Statistics и SAS снижается, а количество ссылок на R –увеличивается [1, Figure 2d]. По результатам изучения реферативной и библиографической базы данных Scopus, произведенным по публикациям всех отраслей знания за 2000 – 2018 гг., ссылки в научных работах на R после 2011 г. преобладают над

ссылками на SPSS [2, Figure 2]. Вместе с тем уточненные результаты поиска по логическим условиям, включающим название соответствующей программы в приставных ссылках научных работ, которые относятся только к области социальных наук, свидетельствуют о большем распространении в практике академических исследований SPSS, чем R (рис. 1).

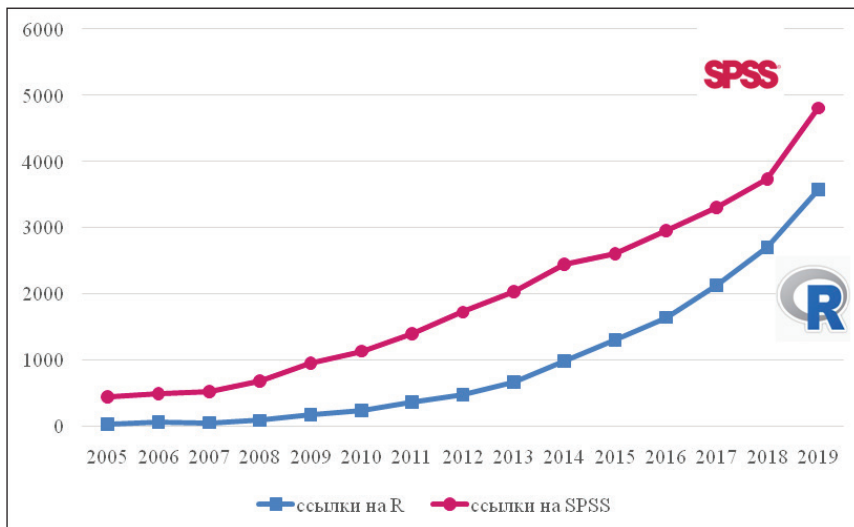


Рис. 1. Динамика ссылок на программы R и SPSS в публикациях, относящихся к области социальных наук, которые индексируются в Scopus (2005–2019 гг.)

Примечание.

Отбор ссылок в базе Scopus осуществлялся 10.03.2020 для временного периода с 2005 по 2019 гг. по таким запросам:

(REF (“the R software”) OR REF (“the R project”) OR REF (“r-project.org”) OR REF (“R development core”)) AND (LIMIT-TO (SUBJAREA, “SOCF”))
REF (SPSS) AND (LIMIT-TO (SUBJAREA, “SOCF”))

Заметим также, что отсутствие возможности вести поиск по ключевым словам, относящимся к названию той или иной программы, в полных текстах научных работ несколько снижает полноту полученных результатов.

Академические исследования всё чаще осуществляются с помощью свободно распространяемого языка программирования и статистического пакета R, что обусловлено наличием большого количества современных методов анализа и визуализации данных, гибкостью, сохранением всей последовательности действий в виде скрипта и пр. Интегрированная среда разработки RStudio, доступ к которой осуществляться как локально, так и удаленно, в облаке, повышает удобство работы с R. Для R создано несколько графических оболочек, не требующих последовательного написания команд при применении основных статистических методов: BlySky Statistics, Deducer, jamovi, JASP, JGR, R Commander, Radiant, Rattle, RKWard.

Учитывая указанные преимущества R, интеграция с ним предусмотрена в распространенном коммерческом (IBM SPSS Statistics, JMP, Microsoft Power BI, RapidMiner, SAS, Stata, Statgraphics, Tableau, TIBCO Statistica, XLSTAT) и бесплатном (KMIME) аналитическом программном обеспечении.

В статистических пакетах предусмотрен импорт и экспорт широкого диапазона форматов данных. Конвертация данных осуществляется с помощью специализированной программы Stat/Transfer.

Важным аспектом исследовательской деятельности является визуальная аналитика и наглядное представление результатов. Для создания научной графики и интерактивных веб-приложений используются языки программирования R (библиотеки flexdashboard, ggplot2, plotly, rgl, scatterplot3d, shiny) и Python (библиотеки bokeh, d3.js, matplotlib, plotly, seaborn), а также программные продукты бизнес-аналитики Microsoft Power BI и Tableau.

Подводя итог, отметим, что современному социологу необходимы навыки работы с офисными приложениями, универсальными статистическими пакетами, а также умения визуализировать данные. Благодаря библиотекам, в которых алгоритмируются новейшие статистические методы, знание вычислительной среды R будет всё более востребовано в передовых академических и прикладных исследованиях.

ЛИТЕРАТУРА

1. Muenchen R. Data Science Software Used in Journals: Stat Packages Declining (including R), AI/ML Software Growing. URL: <http://r4stats.com/2019/04/01/scholarly-datasci-popularity-2019/>.
2. Muenchen R., MacKinnon S. Is Scholarly Use of R Beating SPSS Already? URL: <https://r4stats.com/2019/07/15/is-scholarly-use-of-r-use-beating-spss-already/>.

REFERENCES

1. Muenchen R. (2019) Data Science Software Used in Journals: Stat Packages Declining (including R), AI/ML Software Growing. URL: <http://r4stats.com/2019/04/01/scholarly-datasci-popularity-2019/> [In English].
2. Muenchen R., MacKinnon S. (2019) Is Scholarly Use of R Beating SPSS Already? URL: <https://r4stats.com/2019/07/15/is-scholarly-use-of-r-use-beating-spss-already/> [In English].

Бутенко Марія Костянтинівна

*студентка кафедри маркетингу, менеджменту
та підприємництва Харківського національного університету
імені В. Н. Каразіна*

Петряєв Олексій Олександрович

*кандидат економічних наук, доцент кафедри маркетингу,
менеджменту та підприємництва
Харківського національного університету імені В. Н. Каразіна
ORCID: 0000-0001-6705-4714*

ТЕОРЕТИЧНІ АСПЕКТИ ВИЗНАЧЕННЯ КАДРОВОЇ ПОЛІТИКИ ПІДПРИЄМСТВА

У дослідженні розглянуто основні теоретичні аспекти формування кадрової політики підприємства. Проаналізовано основні види кадрових політик, виявлено їх слабкі та сильні сторони, розглянуто відмінності їх формування. Визначено основні аспекти підбору та формування кадрової політики підприємства.