



Национальный исследовательский университет «Высшая школа экономики»

Факультет социологии

Российское общество социологов

Сообщество профессиональных социологов

СБОРНИК
СТАТЕЙ



НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
**СОВРЕМЕННАЯ СОЦИОЛОГИЯ -
СОВРЕМЕННОЙ РОССИИ**

посвящается памяти
первого декана факультета социологии
**Александра Олеговича
Крыштановского**



МОСКВА
1 – 3
ФЕВРАЛЯ
2012



ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ



Российское общество социологов

Советская социологическая ассоциация - Российское общество социологов



УДК 3160:167/168(06)

ББК 66.5

С56

Редакционный совет

А. Б. Гофман, Г. В. Градосельская, И. Ф. Девятко, Д. Х. Ибрагимова,
И. М. Козина, Л. Я. Косалс, В. А. Мансуров, В. Г. Николаев, О. А. Оберемко,
Н. Е. Покровский, Ю. Н. Толстова, А. Ю. Чепуренко (председатель), Е. Р. Ярская-Смирнова

С56 **Современная социология — современной России: Сборник статей памяти первого декана факультета социологии НИУ ВШЭ А. О. Крыштановского [Электронный ресурс] / НИУ ВШЭ; РОС; СоПСо. — М.: НИУ ВШЭ, 2012. — 753 с. — 1 электрон. диск (CD-ROM).**

ISBN 978-5-904804-08-4

Сборник содержит статьи, подготовленные на основе докладов, сделанных на VI научно-практической конференции «Современная социология — современной России» (г. Москва, 1–3 февраля 2012), посвященной памяти первого декана факультета социологии НИУ ВШЭ Александра Олеговича Крыштановского.

Сборник рассчитан на тех, кто интересуется методологией, методикой и современной практикой проведения социологических исследований.

УДК 3160:167/168(06)

ББК 66.5

ISBN 978-5-904804-08-4



© Национальный исследовательский университет
«Высшая школа экономики», 2012

© Российское общество социологов, 2012

© Сообщество профессиональных социологов, 2012



РАЗДЕЛ 9

«ЕДИНЫЙ АРХИВ СОЦИОЛОГИЧЕСКИХ ДАНЫХ» КАК РЕСУРС ДЛЯ АКАДЕМИЧЕСКИХ ИССЛЕДОВАНИЙ

- А. В. Белецкая**
720 Анализ факторов, определяющих восприятие свободы, с использованием "Единого архива экономических и социологических данных" (ЕАЭСД) НИУ ВШЭ
- А. А. Митюшин**
736 Статистический анализ данных в режиме удаленного доступа: новый сервис "Единого архива экономических и социологических данных" (ЕАЭСД) НИУ ВШЭ



**Анализ факторов,
определяющих восприятие свободы,
с использованием
"Единого архива экономических и
социологических данных" (ЕАЭСД) НИУ ВШЭ**

Анна Владимировна Белецкая

Национальный исследовательский университет
«Высшая школа экономики» (Москва)

В повседневной жизни категория свободы кажется простой и понятной: о ней часто говорят, к ней стремятся, большая степень свободы обычно трактуется как благо. И в политическом лексиконе слово «свобода» употребляется весьма часто, особенно накануне выборов: ее защищают, обещают и гарантируют в случае прихода к власти. Часто можно услышать, что Россия идет по пути создания «свободного общества свободных людей», то есть общества, которое ставит в центр внимания человека, его интересы и проблемы. Однако совпадают ли нужды людей в свободе, представления о ней с условиями для ее реализации, предоставляемыми обществом? Есть ли различия между тем, как понимают свободу граждане, и тем, какой она, согласно представлениям и интересам правящих кругов, реализуется в обществе?

Сложность ответа на вопрос заключается в том, что нет единых представлений о свободе на уровне бытового сознания. Отсутствует также и единое научное определение понятия «свобода». Индивидуальная степень ощущения свободы обусловлена многими факторами, как объективными (ресурсы, которыми обладает человек, условия, в которых он действует, и т.д.), так и субъективными (способность человека выполнять те или иные социальные функции, его место в структурной организации общества и т.д.). Универсальность категории свободы, а также индивидуальный порог возможной и реализуемой свободы формируют у людей субъективное обыденное представление о ней. Кроме того, на субъективное понимание свободы значительным образом влияет специфика культуры, доминирующих в обществе ценностей, идеалов и представлений. В результате интуитивно понятую свободу сложно выразить концептуально. В психологии часто используется тест

Германа Роршаха, смысл которого состоит в том, что предлагают интерпретировать бессмысленную картинку (чернильное пятно), и каждый человек интерпретирует изображение по-своему в зависимости от личного опыта и личностных особенностей, каждый видит то, что наиболее актуально для него. В этом плане понятие «свобода» сильно напоминает тест Г. Роршаха: разные люди, принадлежащие к разным слоям общества, понимают под свободой совершенно разное - это подтверждается и повседневной практикой, и социологическими исследованиями. Если смысловые образы индивидуальной свободы различны, существуют ли факторы, которые оказывают решающее влияние на уровень свободы и, соответственно, на ощущение собственной свободы? Существуют ли различия в том, какое содержание разные социальные группы вкладывают в термин «свобода», какие элементы преобладают в образах индивидуальной свободы (правовые, политические, социально-экономические или другие)? Эти вопросы и по сей день остаются открытыми.

Наше исследование мы начали, отталкиваясь от следующих теоретических посылов. Истинная свобода предполагает не только и не столько «свободу от» (отсутствие притеснений), но и «свободу для» (возможность быть хозяином, творцом своей жизни). Однако, согласно А.Черкасову,¹ в российском обществе свобода чаще всего связывается с разного рода ограничениями. Всевозможные правила и нормы, функционирующие в обществе, оцениваются людьми как главные ограничители свободы, а не как рамки, естественно выработанные человеческим опытом, позволяющие человеку более эффективно реализоваться в обществе. Свобода выступает условием, средством и целью максимальной самореализации человека (цель жизни). Однако, не для каждой личности свобода является первейшей потребностью. Характерными чертами российского человека выступают патерналистское отношение к государству, ожидание от власти благ, а также низкая личная инициатива. «Большинство россиян сегодня ощущают неуверенность, слабую защищенность, иррациональный расчет на первых лиц государства, которые почему-то должны ему помочь»².

В социологическом исследовании проблемы свободы можно выделить две проекции – внутреннюю и внешнюю. Уровень внешней свободы представляет собой возможность и способность субъекта самореализоваться в определенных социальных условиях, не только благодаря этим условиям, но и вопреки различным препятствиям, границам. Внутренний уровень свободы (или психологический аспект проблемы свободы) – это самоощущение социальным субъектом уровня и динамики своей свободы исходя из собственных представлений об актуальных и неактуальных факторах, изменениях социальной среды. Если внешняя свобода зависит не только от субъекта, но и от социальных условий, то внутренняя свобода определяется только субъектом, зависит от его ощущений, ожиданий и аксиологических установок. Люди не могут повлиять на объективные условия своей деятельности, но они обладают индивидуальной степенью свободы при выборе целей и средств их

¹ А. Черкасов «Что такое свобода для российского человека?» / Радио Свободы. Радиogramма. Опубликовано 22.05.2011. <http://www.svobodanews.ru/content/transcript/24178278.html>

² Б. Дубин «Что такое свобода для российского человека?» / Радио Свободы. Радиogramма. Опубликовано 22.05.2011. <http://www.svobodanews.ru/content/transcript/24178278.html>

достижения.. Судить о том, насколько результаты деятельности соответствуют поставленной цели и не противоречат идеалу цели жизни, в полной мере может только сам субъект свободы и деятельности, поэтому внутренняя свобода – это субъективное осознание себя свободным.¹ Ощущение свободы меняет поведение человека, делает его более уверенным в себе, творчески активным. Свобода предполагает несколько уровней существования – уровень потенциальной возможности (как реальной, так и декларируемой) и уровень реализации этих возможностей. Ценность свободы снижается, если нет условий для того, чтобы ею воспользоваться. Но что же в большей степени влияет на ощущение человека себя свободным: условия и средства, которыми располагает человек, или его внутренние субъективные ощущения, представления, оценки? Какой из аспектов - внешняя или внутренняя свобода – оказывается более значимым?

В нашей работе мы используем в качестве индикаторов внешней свободы признаки, определяющие ресурсы и условия деятельности человека: образование, материальную обеспеченность, место проживания, возраст и т.д. Эмпирическими эквивалентами внутренней свободы мы будем считать субъективные оценки респондентом своих обстоятельств, перспектив и возможностей: удовлетворенность жизнью, оценки положения в обществе, возможность изменить жизнь личным усилием и т.д.

Для эмпирической проверки выдвинутых гипотез мы обратились к данным Единого архива экономических и социологических данных (ЕАЭСД) НИУ ВШЭ. В настоящий момент в ЕАЭСД накоплено почти 900 исследований, проведенных лучшими исследовательскими коллективами России. Данные предоставляются для исследовательской и преподавательской работы на безвозмездной основе. Более подробно о деятельности ЕАЭСД можно узнать на сайте: <http://sophist.hse.ru/>

Стоит отметить, что свобода воспринимается подавляющим большинством людей как непреходящая ценность. В частности, это подтверждают результаты ряда исследований,² в которых респонденты отвечали на вопросы о самых значимых ценностях и своих ассоциациях с понятием «свобода». И в 80-х, и в 90-х, и в 2000-х годах порядка 90% опрошенных определяли свободу как самую важную ценность, вызывающую положительные эмоции.

¹ Белецкая А.В. Социально-философские основания свободы // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. - Тамбов: Грамота, 2011. № 5 (11): в 3-х ч. Ч. I. С. 26-30.

² Исследование «Ценности-84: сравнительный анализ интериоризации ценностей в развитом социалистическом обществе», проведено Институтом социологии АН СССР в 1984 году (1262 опрошенных, выборка территориальная, четырехступенчатая). Исследование «Наши ценности и интересы сегодня» проведено Центром социального прогнозирования и маркетинга под рук. Ф.Э. Шереги в 1998 году (1100 опрошенных, Опрос проводился в 12 территориально-экономических районов и г. Москва). Исследование «Россияне о судьбах России в XX веке и своих надеждах на XXI век» проведено Центром социального прогнозирования и маркетинга в 2000 году (2054 опрошенных, выборка квотная, случайная). Исследования размещены на сайте ЕАЭСД: <http://sophist.hse.ru/db/oprosy.shtml?ts=2> .

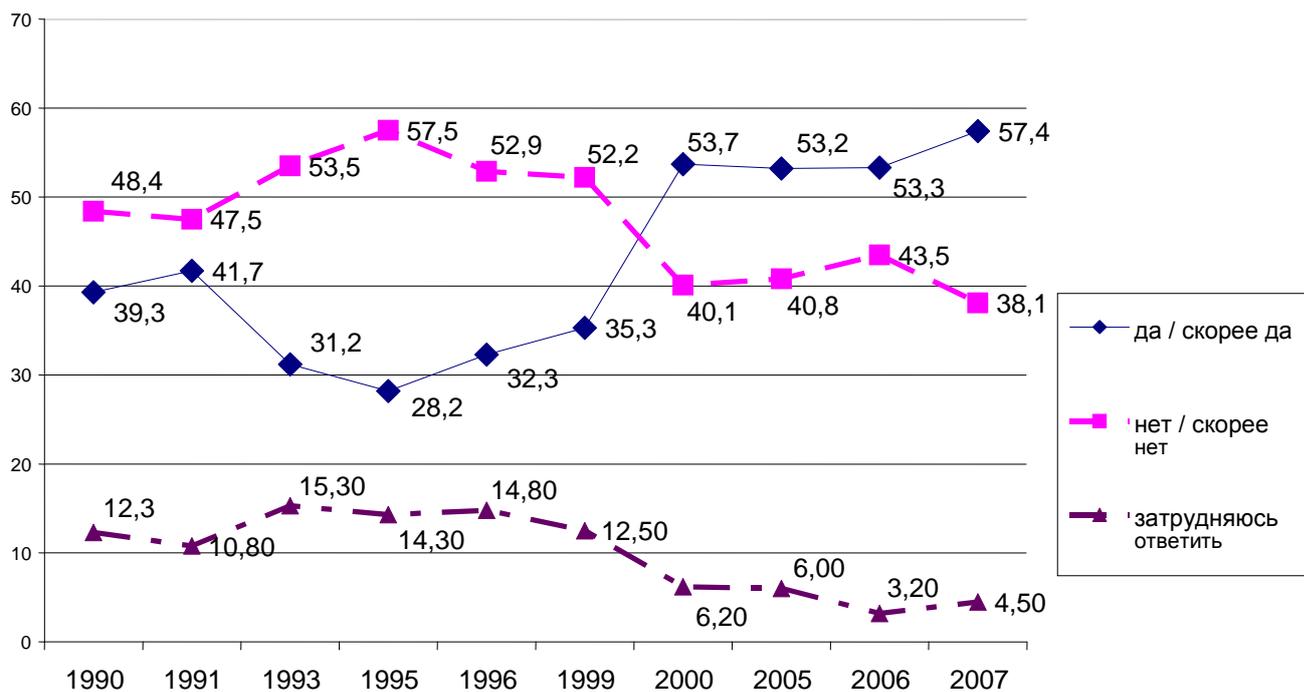


Рисунок 1. Чувствуете ли Вы себя в нашем обществе свободным человеком?

На Рис.1 представлена динамика ответов респондентов на вопрос «Чувствуете ли вы себя свободным человеком?»¹ Начиная с 1991 и вплоть до 1995 года, доля респондентов, считающих себя свободными, уменьшилась с 42% до 28%, в 1995 году начался некоторый рост показателя. Резкое увеличение числа «свободных» людей произошло в 1999-2000 годах, далее показатель почти не менялся, лишь немного увеличился в 2007 году. Отметим переломный характер 2000 года еще в одном отношении: доля респондентов, ощущающих собственную свободу (54%), впервые превысила количество «несвободных» (40%). Обратим внимание на то, что динамика распределения ответов «затрудняюсь ответить» практически повторяет динамику ответов «не чувствую себя свободным»: в те года, когда доля «несвободных» возрастала, росла и доля сомневающихся, и наоборот. Напротив, в те периоды, когда росло число «свободных», уменьшалась доля неуверенных. Самый низкий процент (3%) респондентов, затруднившихся с ответом относительно ощущения собственной свободы, был зафиксирован в 2006 году, самый высокий (15%) - в 1993.

Какие факторы в большей степени влияют на такое распределение ответов респондентов? В какой мере изменения в ощущениях респондентами личной свободы, представленные на Рис.1, являются следствием обстановки в обществе, в частности, в области экономики,

¹ Все исследования проведены по всероссийской многоступенчатой стратифицированной случайной выборке: исследования «Омнибус» проведены ВЦИОМ в 1990 (3198 опрошенных) и 1991 (3061 опрошенных) годах, «Мониторинг социально-экономических перемен» проведены ВЦИОМ в 1993 (2018 опрошенных), 1995 (1989 опрошенных), 1996 (2426 опрошенных), 1999 (2385 опрошенных) годах; «Курьер» проведены АНО «Левада - центр» в 2000 (1600 опрошенных), 2005 (1584 опрошенных), 2006 (1610 опрошенных) и 2007 (1600 опрошенных) годах. Исследования размещены на сайте ЕАЭСД: <http://sophist.hse.ru/db/oprosy.shtml?ts=2>

политики и т.д.? Является ли эта динамика ответной реакцией на изменения индивидуальной ситуации респондента, своего рода мерой его относительной депривации? На Рис.2 мы совместили несколько графиков – ответы на вопрос о самых актуальных проблемах в нашем обществе и долю тех, кто ощущает себя свободным человеком.

В 1991 году отмечается снижение остроты социальных проблем: доля тех, кто выбирает рост цен в качестве актуальной проблемы уменьшается с 66% до 60%,; рост преступности - с 41% до 10%; взяточничество - с 40% до 10%. В это же время, пусть незначительно, но повышается ощущение респондентами собственной свободы (с 39% до 42%). Период 1995-1996 годов характеризуется пиком остроты проблем (82%, 65%, 26%) и тут же отмечается рекордное падение (28%) доли тех, кто ощущает себя свободным. В отмеченный ранее «переломный» 2000 год резко уменьшается острота выбранных проблем: до 50% снизилась доля респондентов, обеспокоенных повышением цен, что является самым низким показателем за исследуемый интервал времени; также только 16% и 18% респондентов волнуют проблемы роста преступности и взяточничества соответственно, и, как мы видим, опросы фиксируют небывалый ранее «взлет» свободы (до 54%). 2005 год несколько выпадает из общей логики, так как очередное повышение остроты проблем (до 75%, 32%, 27%) не оказывает влияния на ощущение свободы респондентами, более того, в промежутке с 2000 по 2006 года восприятие индивидуальной свободы остается стабильным.

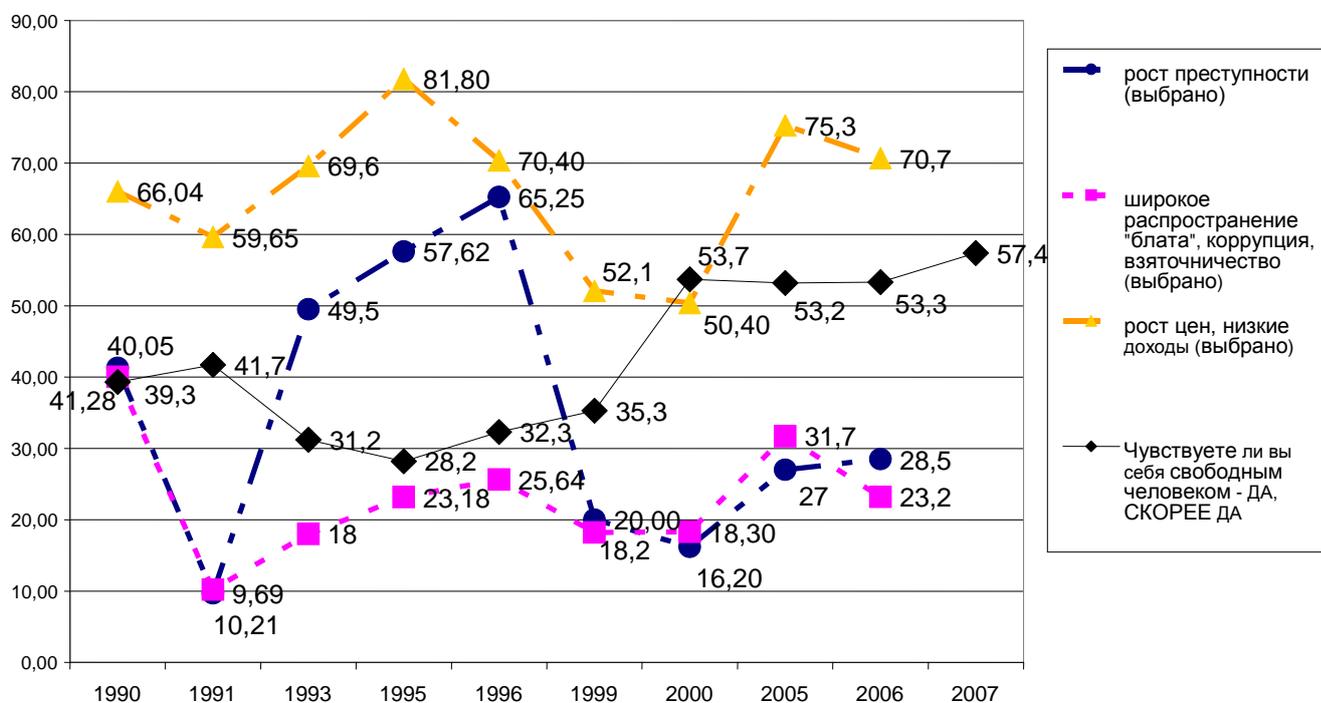


Рисунок 2. Какие проблемы нашего общества, на Ваш взгляд, являются наиболее острыми? / Чувствуете ли Вы себя свободным человеком?

Данные, представленные на Рис.2, фиксируют наличие связи между одним из эмпирических индикаторов внешней свободы (остротой давящих проблем) и ощущением индивидуальной свободы. Для дальнейшего анализа и выявления факторов, влияющих на восприятие личной свободы, мы выбрали данные 2005 года¹, поскольку этот период характеризуется стабильностью оценок.

В качестве основной процедуры дальнейшего анализа была выбрана логистическая регрессия, позволяющая непосредственно оценить вероятность наступления события – в нашем случае таким событием является субъективное ощущение респондентом себя свободным. В качестве независимых предикторов мы выбрали два блока факторов, определяющих внутренний и внешний уровни свободы.² К сожалению, индикаторы, важные, с нашей точки зрения, для анализа, обладают высокой степенью корреляции и не могут быть непосредственно включены в регрессионную модель в качестве независимых переменных. Нам пришлось выполнить некоторые предварительные процедуры, чтобы демпфировать эффект мультиколлинеарности.

Мы предположили, что наличие высшего образования обостряет и делает более осознанным чувство свободы респондента. Образно говоря, человек становится более «капризным» в своем отношении к свободе: люди, владеющими рефлексивными процедурами и обладающие более широким кругозором, могут острее ощущать давление обстоятельств, что снижает ощущение «я - свободен». Данные, приведенные в Таб.1, показывают, что, действительно, по мере увеличения уровня образования уменьшается ощущение чувства свободы: от 68% «свободных» при образовании «начальное и ниже» до 54% в случае наличия у респондента высшего образования. Так как для нашей модели мы отбираем показатели, которые положительно влияют на ощущение свободы, было бы логичным взять альтернативу «образование начальное или ниже», однако низкая неравномерная наполненность граф таблицы, в частности по данной альтернативе (4,5%), не позволяет нам этого сделать.

¹Исследование «Курьер–2005» (время проведения: 22.01.2005 - 01.02.2005, число опрошенных: 1584. Полевые работы произвел АНО «Левада - центр») размещено на сайте ЕАЭСД: <http://sophist.hse.ru/db/oprosy.shtml?ts=2>

² Отбор индикаторов, характеризующих уровни свободы, был ограничен возможностями вторичного анализа и осуществлялся, исходя из возможностей, предоставляемых анкетой выбранного исследования.

Таблица сопряженности: ваше образование?
/ чувствуете ли вы себя в нашем обществе свободным человеком?

		ЧУВСТВУЕТЕ ЛИ ВЫ СЕБЯ В НАШЕМ ОБЩЕСТВЕ СВОБОДНЫМ ЧЕЛОВЕКОМ?		Итого
		0 не чувствую себя свободным	1 чувствую себя свободным	
ВАШЕ ОБРАЗОВАНИЕ?	1 начальное или ниже (не закончена неполная средняя школа)	20 31,7%	43 68,3%	63 100,0%
	2 неполная средняя школа (7-8, сейчас 9 классов)	112 43,2%	147 56,8%	259 100,0%
	3 ПТУ на базе неполной средней школы	28 34,1%	54 65,9%	82 100,0%
	4 средняя школа (10, сейчас 11 классов)	123 41,4%	174 58,6%	297 100,0%
	5 ПТУ на базе средней школы	52 41,9%	72 58,1%	124 100,0%
	6 среднее специальное образование (техникум)	205 45,2%	249 54,8%	454 100,0%
	7 незаконченное высшее (не менее 3-х курсов вуза)	20 45,5%	24 54,5%	44 100,0%
	8 высшее (закончен вуз)	104 45,6%	124 54,4%	228 100,0%
Итого		664 42,8%	887 57,2%	1551 100,0%

Изначально мы предполагали, что респонденты, «обремененные» семьей и детьми, теряют чувство свободы, у них оно вытесняется чувством ответственности и заботы. Простой анализ таблицы сопряженности показывает, что наличие или отсутствие детей у респондента никак не отражается на его оценке собственной свободы (Таб.2), чего нельзя сказать о семейном положении. В Таб.3 приведены результаты логистической регрессии (семейное положение было включено в модель как категориальная переменная). Только позиция «холост/незамужем и никогда не состоял(а) браке» обладает достаточной значимостью для нашего анализа.

Таблица 2

Таблица сопряженности: чувствуете ли вы себя в нашем обществе свободным человеком? * есть дети в возрасте до 16 лет?

		ЕСТЬ ДЕТИ В ВОЗРАСТЕ ДО 16 ЛЕТ		Итого
		0 нет детей в возрасте до 16 лет	1 есть дети в возрасте до 16 лет	
ЧУВСТВУЕТЕ ЛИ ВЫ СЕБЯ В НАШЕМ ОБЩЕСТВЕ СВОБОДНЫМ ЧЕЛОВЕКОМ?	0 не чувствую себя свободным	338 42,6%	253 43,0%	591 42,8%
	1 чувствую себя свободным	456 57,4%	335 57,0%	791 57,2%
Итого		794 100,0%	588 100,0%	1382 100,0%

Таблица 3

Таблица коэффициентов логистической регрессии
(зависимая переменная - Чувствуете ли вы себя в нашем обществе свободным человеком? / независимая переменная - Каково ваше семейное положение? (включена в модель как категориальная))

		B	Стд.Ошибка	Вальд	ст.св.	Знч.	Exp(B)
Шаг 1 D5. КАКОВО ВАШЕ СЕМЕЙНОЕ ПОЛОЖЕНИЕ?	1 холост/незамужем и никогда не состоял(а) в браке	,446	,204	4,755	1	,029	1,562
	2 живем вместе, не оформляя отношений официально	,423	,278	2,306	1	,129	1,526
	3 замужем/женат, но живем раздельно	-,313	,249	1,582	1	,208	,731
	4 замужем/женат и живем вместе	,071	,167	,180	1	,672	1,073
	5 разведен(а)	-,289	,228	1,598	1	,206	,749
	Константа	,216	,151	2,032	1	,154	1,241

Было бы естественно ожидать, что наличие материальных проблем у человека негативно отражается на ощущении свободы: постоянная забота об удовлетворении первостепенных материальных потребностей, казалось бы, должна быть дополнительным гнѐтом. Однако, согласно данным, представленным в Таб.4, среднемесячный доход в группе респондентов, ощущающих собственную свободу, незначительно отличается от среднего дохода в группе респондентов, считающих себя несвободными. Более того, T-test показывает, что мы не можем считать эти различия значимыми (Таб.5).

Таблица 4

Зависимость между чувством личной свободы и
среднемесячным доходом респондента

чувствуете ли вы себя в нашем обществе свободным человеком?	Среднемесячный доход	N	Std. Deviation	% of Total N	% of Total Sum
0 не чувствую себя свободным	2866,47	618	2446,582	58,2%	60,7%
1 чувствую себя свободным	2581,97	444	1849,085	41,8%	39,3%
Итого	2747,53	1062	2219,950	100,0%	100,0%

T-test: Зависимость между чувством личной свободы и среднемесячным доходом респондента

	Levene's Test for Equality of Variances		t-test for Equality of Means							
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	
qv2 СРЕДНЕ-МЕСЯЧНЫЙ ДОХОД	Equal variances assumed	3,159	,076	,987	365	,324	233,629	236,727	231,892	699,149
	Equal variances not assumed			1,034	364,989	,302	233,629	225,951	210,701	677,958

Не менее важным фактором, влияющим на ощущение себя свободным, с нашей точки зрения, является возраст. Мы предположили, что более молодые респонденты чувствуют себя более свободными, а с возрастом чувство свободы уменьшается. На Рис.3 представлена динамика показателя «свободы» в зависимости от возраста. Чувство свободы максимально в молодых возрастах и начинает заметно снижаться при приближении к тридцатилетнему рубежу. В возрасте примерно 45 лет доли «свободных» и «несвободных» сравниваются и в старших возрастах ощущение себя «несвободным» доминирует. Резкая перемена наступает после 70 лет.

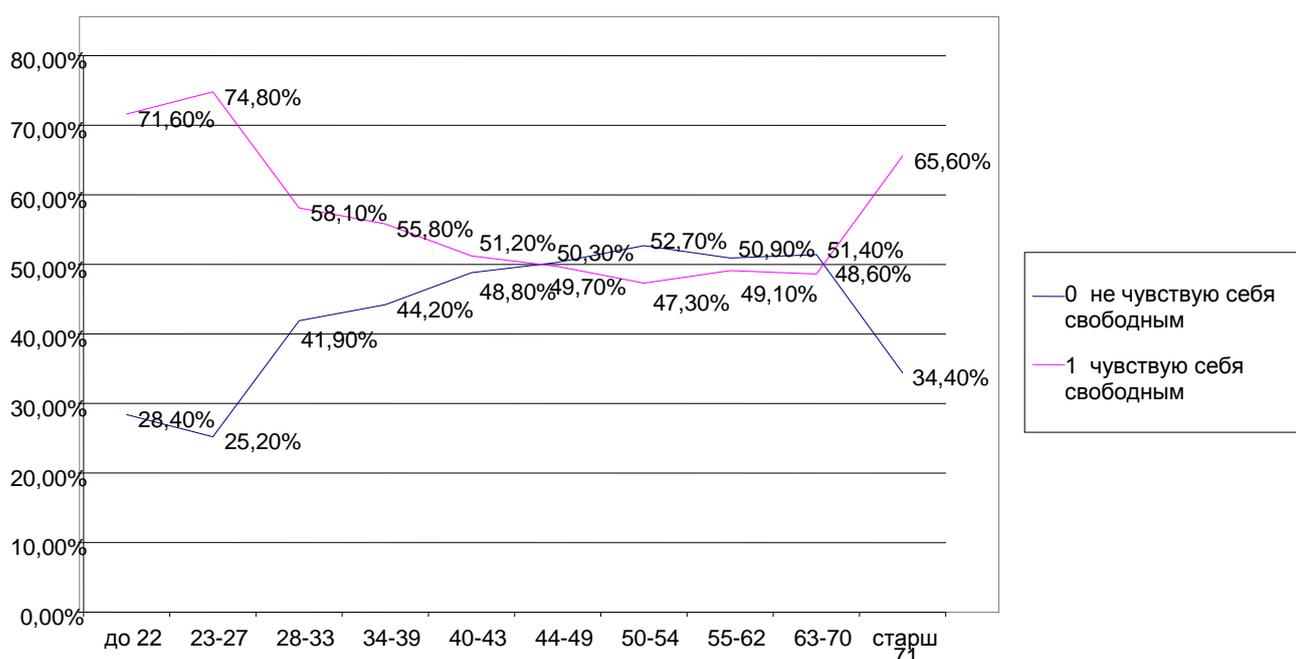


Рисунок 3. Зависимость между чувством личной свободы и возрастом респондента

Было бы естественным предположить, что на ощущение свободы влияет место проживания: с увеличением размера населенного пункта, в котором живет человек, растет и количество возможностей, которые у него появляются, и способов их реализации. Однако мы установили, что это верно лишь отчасти. Так, значимое влияние на ощущение свободы оказывает не проживание в мегаполисах, как было бы логично ожидать, а в «городах с населением более 500 тыс.», при этом проживание в меньших населенных пунктах не оказывает сколько-нибудь значимого влияния. В Таб.6 приведены результаты логистической регрессии (место проживания респондента было включено в модель как категориальная переменная), позволяющей судить о том, что только позиция «город с населением более 500 тыс.» обладает достаточной значимостью для нашего анализа.

Таблица 6

**Таблица коэффициентов логистической регрессии
(зависимая переменная - Чувствуете ли вы себя
в нашем обществе свободным человеком?
/ независимая переменная – Населенный пункт,
в котором проживает респондент
(включена в модель как категориальная))**

		В	Стд.Ошибка	Вальд	ст.св.	Знч.	Exp(B)
Шаг 1^a	Населенный пункт			18,172	4	,001	
	Москва	-,305	,195	2,463	1	,117	,737
	более 500 тыс.	-,372	,146	6,474	1	,011	,690
	от 100 до 500 тыс.	,055	,150	,132	1	,716	1,056
	от 10 до 100 тыс	,246	,149	2,727	1	,099	1,279
	Константа	,334	,094	12,632	1	,000	1,397

С нашей точки зрения, было бы важно включить в анализ степень коррумпированности социальной среды как показатель свободы реализации намерений и планов. В качестве такого индикатора был выбран вопрос о том, давал ли респондент за последний год взятки (от взяток при приеме ребенка в детский сад, до взяток за оформление документов). Мы выбрали категорию «не давал взятку» как показатель, определяющий внешний уровень свободы. Все показатели, отобранные нами в качестве индикаторов внешнего и внутреннего уровней свободы, приведены в Таб.7

**Индикаторы внешнего и внутреннего уровня свободы,
обобранные в результате предварительного анализа**

	<i>Вопросы анкеты</i>	<i>Альтернативы, обладающие наибольшим влиянием на оценку степени личной свободы в целом</i>
Внешний уровень свободы	1. Возраст респондента	«моложе 40 лет»
	2. Семейное положение респондента	«холост/незамужем и никогда не состоял(а) в браке»
	3. Размер населенного пункта, в котором проживает	«город с населением более 500 тыс.»
	4. Степень коррумпированности социальной среды: вопрос «Давали ли Вы взятки за последний год?»	«не давал взятки за последний год»
Внутренний уровень свободы	1. Насколько Вы в настоящее время удовлетворены своим положением в обществе	«вполне удовлетворен/скорее удовлетворен»
	2. Насколько Вы в настоящее время удовлетворены своим образованием	«вполне удовлетворен/скорее удовлетворен»
	3. Насколько Вы в настоящее время удовлетворены своим материальным положением	«вполне удовлетворен/скорее удовлетворен»
	4. Как изменятся в течение следующих шести месяцев позиции России на мировой арене улучшатся?	«улучшатся»
	5. Как изменится качество жизни респондента в течение следующих шести месяцев, по сравнению с нынешним?	«улучшится / останется таким же как сейчас»
	6. Чувствуете ли Вы уверенность в завтрашнем дне?	«да/скорее да»
	7. Локус контроля: от чего в большей степени зависит благополучие человека (от самого человека или от того, насколько справедливо устроено общество)?	«благополучие зависит от самого человека»

Для того чтобы избежать явления мультиколлинеарности, но учесть влияние всех выделенных нами факторов, были построены индексы внешней и внутренней свободы по следующему алгоритму: респондент получал один балл, если обладал «шансом на свободу» по каждому из отобранных выше факторов. Так, если возраст респондента был до 40 лет, то его индекс внешней свободы увеличивался на единицу. «Шансы на свободу» суммировались. Таким образом, индекс внешней свободы, включающий 4 фактора, укладывается в интервал от 0 до 4, индекс внутренней свободы, состоящий из 7 факторов, находится в диапазоне от 0 до 7. Распределение полученных индексов приведено в Таб.8 и Таб.9.

Таблица 8

Внешний уровень свободы (индекс)

	Частота	Процент	Валидный процент	Кумулятивный процент
Валидные 0	153	9,3	9,3	9,3
1	729	44,2	44,2	53,4
2	532	32,2	32,2	85,6
3	206	12,5	12,5	98,1
4	31	1,9	1,9	100,0
Итого	1651	100,0	100,0	

Таблица 9

Внутренний уровень свободы (индекс)

	Частота	Процент	Валидный процент	Кумулятивный процент
Валидные 0	88	5,3	5,3	5,3
1	205	12,4	12,4	17,7
2	319	19,3	19,3	37,1
3	329	19,9	19,9	57,0
4	305	18,5	18,5	75,5
5	224	13,6	13,6	89,0
6	153	9,3	9,3	98,3
7	28	1,7	1,7	100,0
Итого	1651	100,0	100,0	

Поскольку коэффициент корреляции между построенными индексами внутренней и внешней свободы низкий (см. Таб.10), мы можем включить их в нашу регрессионную модель как независимые факторы. Полученные результаты представлены в Таб.11, Таб.12 и Таб.13.

Таблица 10

Корреляции индексов внешнего и внутреннего уровне свободы

		внешний уровень свободы (индекс)	внутренний уровень свободы (индекс)
внешний уровень свободы (индекс)	Pearson Correlation	1	,197(**)
	Sig. (2-tailed)		,000
	N	1651	1651
внутренний уровень свободы (индекс)	Pearson Correlation	,197(**)	1
	Sig. (2-tailed)	,000	
	N	1651	1651

Таблица 11

Таблица классификаций логистической регрессии
(зависимая переменная: чувствуете ли Вы себя в нашем обществе
свободным человеком? ковариаты: внешний уровень свободы
(индекс) и внутренний уровень свободы (индекс))

	Наблюденные		Предсказанные		
			чувствую себя свободным		Процент корректных
			не чувствую себя свободным	чувствую себя свободным	
Шаг 1	чувствую себя свободным	не чувствую себя свободным	369	295	55,6
		чувствую себя свободным	189	698	78,7
	Общий процент				68,8

Уровень предсказательной силы модели, которую мы получили, составляет 69%. Однако только индекс внутренней свободы обладает необходимым уровнем значимости. Индекс же внешней свободы не является значимым и на втором шаге исключается из модели (Таб.12).

Таблица 12

Переменные в уравнении
(зависимая переменная: чувствуете ли Вы себя в нашем обществе
свободным человеком? ковариаты: внешний уровень свободы (индекс)
и внутренний уровень свободы (индекс))

		B	Стд.Ошибка	Вальд	ст.св.	Знч.	Exp(B)
Шаг 1(a)	внешний уровень свободы (индекс)	-,023	,065	,129	1	,719	,977
	внутренний уровень свободы (индекс)	,586	,038	232,739	1	,000	1,798
	Константа	-1,502	,148	102,595	1	,000	,223
Шаг 2(a)	внутренний уровень свободы (индекс)	,584	,038	239,610	1	,000	1,793
	Константа	-1,529	,127	145,023	1	,000	,217

Таблица 13

Переменные, не включенные в уравнение
(зависимая переменная: чувствуете ли Вы себя в нашем обществе
свободным человеком? ковариаты: внешний уровень свободы (индекс)
и внутренний уровень свободы (индекс))

		Балл	ст.св.	Знч.
Шаг 2(a)	Переменные внешний уровень свободы (индекс)	,129	1	,719
	Обобщенные статистики	,129	1	,719

Построенная нами модель подтверждает наличие влияния внутренней свободы на ощущение индивидуальной свободы в целом. Но не менее интересно проверить, какие факторы, входящие в индекс внутренней свободы, обладают большим влиянием на оценку респондентом степени личной свободы. Поскольку коэффициенты корреляции между факторами, включенными в индекс внутренней свободы, не высоки (Таб.14), мы включили их в нашу регрессионную модель как независимые предикторы. Полученные результаты представлены в Таб.15 и Таб.16.

Таблица 14

Корреляции факторов, включенных в индекс внутренней свободы¹

		Удовлетворенность своим положением в обществе	Удовлетворенность образованием	Удовлетворенность материальным положением	Положение России на мировой арене улучшится	Качество жизни респондента улучшится или останется таким же	Уверенность респондента в завтрашнем дне	Благополучие человека зависит от него самого
удовлетворенность своим положением в обществе	Pearson Correlation	1	,205	,239	,058*	,177	,250	,170
	Sig. (2-tailed)		,000	,000	,029	,000	,000	,000
удовлетворенность образованием	Pearson Correlation	,205	1	,170	,004	,069	,102	,066
	Sig. (2-tailed)	,000		,000	,868	,008	,000	,009
удовлетворенность материальным положением	Pearson Correlation	,239	,170	1	,039	,180	,318	,183
	Sig. (2-tailed)	,000	,000		,135	,000	,000	,000
Положение России на мировой арене улучшится	Pearson Correlation	,058*	,004	,039	1	,156	,109	,103
	Sig. (2-tailed)	,029	,868	,135		,000	,000	,000
качество жизни респондента улучшится или останется таким же	Pearson Correlation	,177	,069	,180	,156	1	,287	,166
	Sig. (2-tailed)	,000	,008	,000	,000		,000	,000
уверенность респондента в завтрашнем дне	Pearson Correlation	,250	,102	,318	,109	,287	1	,359
	Sig. (2-tailed)	,000	,000	,000	,000	,000		,000
благополучие человека зависит от него самого	Pearson Correlation	,170	,066	,183	,103	,166	,359	1
	Sig. (2-tailed)	,000	,009	,000	,000	,000	,000	

¹ Рассчитано для 1651 респондентов.

Таблица 15

Переменные в уравнении (зависимая переменная: чувствуете ли Вы себя в нашем обществе свободным человеком?
ковариаты: факторы, составляющие индекс внутренней свободы)

		В	Стд.Ошибка	Вальд	ст.св.	Знч.	Exp(B)
Шаг 2(a)	уверен в завтрашнем дне	1,661	,156	113,785	1	,000	5,263
	благополучие человека зависит главным образом от него самого	,737	,143	26,745	1	,000	2,090
	удовлетворен своим положением в обществе	,403	,146	7,637	1	,006	1,497
	удовлетворен своим образованием	,297	,140	4,501	1	,034	1,346
	удовлетворен своим материальным положением	,497	,156	10,130	1	,001	1,645
	положение России на мировой арене в ближайшие шесть месяцев улучшится	,427	,179	5,684	1	,017	1,533
	Константа	-1,251	,147	72,442	1	,000	,286

Таблица 16

Переменные, не включенные в уравнение (зависимая переменная: чувствуете ли Вы себя в нашем обществе свободным человеком? ковариаты: факторы, составляющие индекс внутренней свободы)

Шаг 2(a)	Переменные	качество жизни респондента в ближайшее время улучшится, или, по крайней мере, останется таким же	Балл	ст.св.	Знч.
			,524	1	,469
	Обобщенные статистики		,524	1	,469

Данные, приведенные в Таб.15, показывают, что влияние факторов, составляющих индекс внутренней свободы, различно. Так, уверенность респондента в завтрашнем дне оказывает самое сильное влияние на ощущение себя свободным. При фиксировании прочих переменных полученные значения коэффициентов свидетельствуют о том, что уверенность в завтрашнем дне увеличивает отношение шансов «свободен» и «несвободен» в 5,2 раз; уверенность в том, что благополучие человека зависит главным образом от него самого – в 2 раза; удовлетворенность своим материальным положением – 1,6 раз; мнение, что положение России на мировой арене в ближайшие шесть месяцев улучшится – в 1,5 раз; удовлетворенность своим положением в обществе – в 1,5 раз; удовлетворенность своим образованием – в 1,3 раз. Представления респондента о том, что качество его жизни в ближайшее время улучшится, или, по крайней мере, останется таким же, не обладает достаточным уровнем значимости.

Наше предположение о большей значимости внутренних ощущений по сравнению с внешними обстоятельствами нашло подтверждение в ходе проведенного анализа. Значимыми в построенной модели логистической регрессии выступают только индикаторы внутренней свободы социального субъекта (оценки, степень удовлетворенности, ощущения). Начиная исследование, мы отталкивались от определения свободы как способности субъекта к гармоничному (когда возможности и способности человека соответствуют поставленным целям) осуществлению деятельности, разворачивающейся в границах, предоставляемых ему условиями окружающей действительности, с целью наиболее полной самореализации. Мы пришли к выводу, что человек, действительно, ощущает себя свободным, если в своем стремлении реализоваться в обществе полагается на себя, а его возможности и способности находятся в состоянии гармонии (человеку достаточно того, чем он обладает для реализации своих целей) и он при этом чувствует уверенность в завтрашнем дне. Все остальное – имеет меньшую значимость. На ощущение собственной свободы в большей степени влияет свобода внутренняя, внешняя проекция уходит на второй план, выступая полем для реализации потенциальных возможностей социального субъекта.



Статистический анализ данных в режиме удаленного доступа: новый сервис "Единого архива экономических и социологических данных" (ЕАЭСД) НИУ ВШЭ

Антон Александрович Митюшин

**Национальный исследовательский университет
«Высшая школа экономики» (Москва)**

1. Введение

Международное научное сообщество давно осознало факт необходимости сохранения опросных данных и организации доступа к ним для вторичного анализа. Первый архив данных по социальным наукам был создан еще в 1946 году – это знаменитый Ропер-центр.¹ Значительные шаги в этом направлении были предприняты в конце 1950-х — 1960-х, когда возникло большое количество американских и европейских общедоступных архивов данных. В России аналогичная деятельность была начата лишь спустя несколько десятилетий².

На данный момент в архивах данных по социальным наукам накоплены сотни тысяч исследований, тематически охватывающих значительное количество проблем, многие из которых все еще ждут своего исследователя. Свое место в общем ряду нашел и Единый архив экономических и социологических данных (ЕАЭСД) НИУ ВШЭ: http://sophist.hse.ru/arch_about.shtml. В настоящий момент коллекции ЕАЭСД насчитывают порядка 900 опросов, проведенных лучшими российскими исследовательскими институтами в период с 1970-х гг. по настоящее время. Кроме того, ЕАЭСД предоставляет удобные, постоянно совершенствующие средства поиска и статистического анализа данных в режиме удаленного доступа. Все это отражается на показателях успешной деятельности ЕАЭСД: в настоящий момент ЕАЭСД входит в число 20 популярнейших ресурсов Рунета по социологии³. Особо следует отметить, что ЕАЭСД занял достойное место в международном сообществе архивов данных по социальным наукам, с одной стороны, активно продвигая накопленные коллекции в зарубежное исследовательское сообщество и, с другой стороны, внедряя международные стандарты в свою работу.

¹ <http://www.ropercenter.uconn.edu/>

² См. подробнее Scheuch, E. K. (2003), History and visions in the development of data services for the social sciences. *International Social Science Journal*, vol. 55, Issue 177, p. 385–399. Ralph L. Bisco. Social science data archives: progress and prospects // *Social Science Information* February 1967 vol. 6 no. 1, p. 39-40; Scheuch, E. K. (2003), History and visions in the development of data services for the social sciences. *International Social Science Journal*, vol. 55, Issue 177, p. 385–399; Хахулина Л.А., Косова Л.Б. Создание единого архива социологических данных: проблемы и перспективы // СПЕРО, № 6, С. 203-208 М., 2007.

³ По данным рейтинга Openstat <http://rating.openstat.ru/?cid=2117188>

Одним из сервисов, внедренных ЕАЭСД в свою практику, стала возможность анализа данных в режиме удаленного доступа, без необходимости загрузки данных на компьютер пользователя, что обеспечивает новые возможности для преподавания социальных дисциплин и исследовательской работы. Не осуществляя предварительного заказа данных и не тратя ресурсы на приобретение программ для статистического анализа пользователи ЕАЭСД получают широкий спектр возможностей работы с данными.

В 2005 г. ЕАЭСД первым в России освоил и установил на своем сайте NESSTAR — программный комплекс, разработанный Архивом данных Великобритании (UK Data Archive) и Норвежким службой данных по социальным наукам (Norwegian Social Science Data Services) специально для архивов социологических данных. Этот программный продукт основан на новом поколении веб-технологий и обеспечивает возможность статистической обработки данных прямо на сайте. К сожалению, несмотря на уверения разработчиков в том, что в NESSTAR будет реализовано большое количество функций и он станет поистине «Dream machine»¹ для социальных наук, этого не произошло - за последние 7 лет набор доступных функций практически не изменился. В связи с этим ЕАЭСД обратил внимание на динамично развивающийся и весьма перспективный программный продукт Survey and Documentation Data Analysis (SDA) разработанный Калифорнийским университетом в Беркли. Опыт использования этого продукта одним из ведущих мировых архивов ICPSR², дает дополнительные основания обратить внимание на этот программный комплекс.

Характерными особенностями системы SDA являются³:

- возможность подготовки переменных для анализа (перекодировка, вычисления, работа с пропущенными значениями, фильтры);
- возможность статистического анализа (построение одномерных линейных распределений и таблиц сопряженности, сравнение средних и дисперсионный анализ, проведение анализа надежности, расчет различных коэффициентов корреляции и построение корреляционных матриц. Отдельно выделим возможности регрессионного моделирования различных видов: линейная регрессия, логистическая, пробит).
- возможность расчета комплексных стандартных ошибок. Дело в том, что при анализе случайных и квотных выборок должны использоваться различные процедуры для расчета стандартных ошибок и доверительных интервалов. SDA обеспечивает возможность использования этих процедур при расчете процентов, средних значений, различий между средними значениями и регрессионных коэффициентов.
- возможность построения графиков нескольких типов: столбчатых, штабельных, линейных и секторных диаграмм при работе с некоторыми процедурами.
- быстрое получение результатов вне зависимости от размера выборки (пакет позволяет обрабатывать тысячи переменных и миллионы случаев).¹

1 Jostein Ryssevik and Simon Musgrave. The Social Science Dream Machine: Resource Discovery, Analysis, and Delivery on the Web// *Social Science Computer Review* Summer 2001 vol. 19 no. 2 163-174.

² <http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/sda.jsp>

³ <http://sda.berkeley.edu/>

Добавим, что программный комплекс достаточно прост в освоении, так как обладает интуитивно понятным интерфейсом. Не последнюю роль в этом играет качественная, хорошо продуманная контекстная справочная система², использование которой помогает «не заблудиться» в функциях и процедурах программного комплекса. Конечно, это не означает, что система вообще лишена какого-рода недостатков - мы обратим на них внимание далее. Но достоинства, безусловно, перевешивают. Остановимся подробнее на некоторых особенностях работы в программном комплексе SDA.

2. Основные особенности работы с программным комплексом SDA

2.1. Начало работы

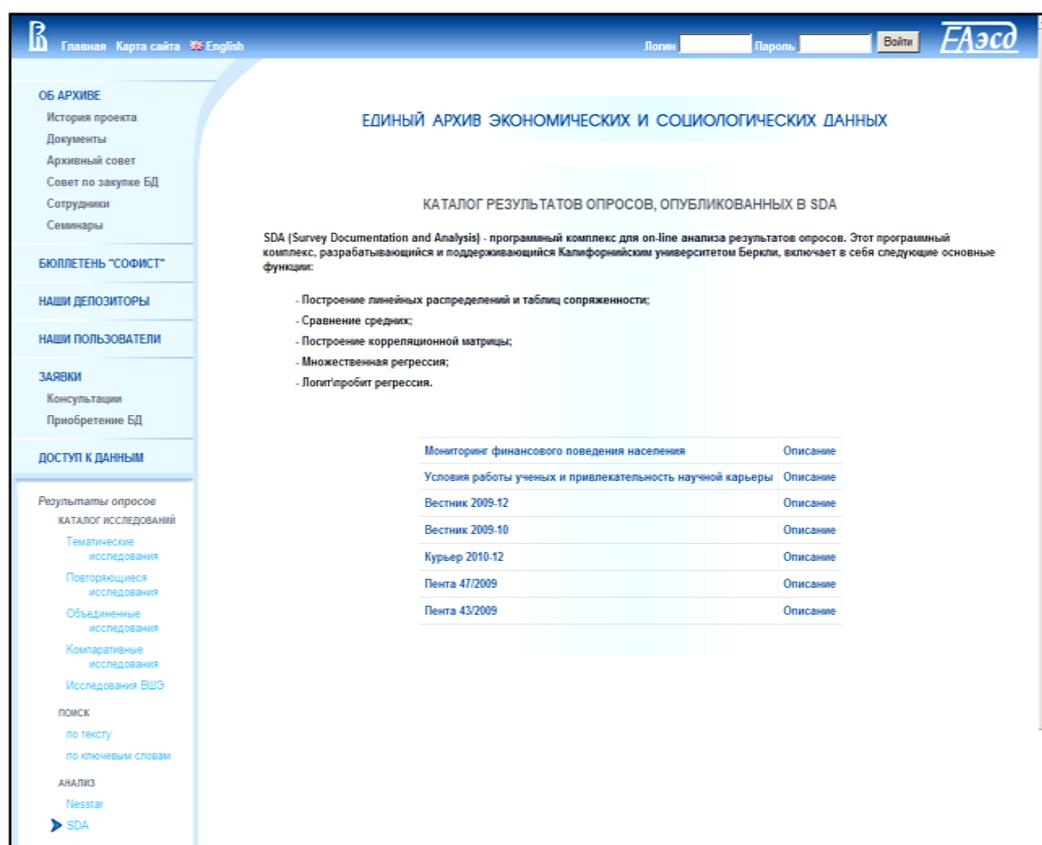


Рисунок 1. Каталог исследований опубликованный в SDA

По своей структуре и подходу к работе с данными интерфейс SDA очень напоминает интерфейс Nesstar, о котором мы писали выше, поэтому пользователю, уже имевшему опыт работы с Nesstar, будет легко переключиться на работу с SDA. Для того, чтобы войти систему SDA³ необходимо в левой панели меню сайта EAЭСД в подразделе «Результаты

¹ Использование нами SDA на практике показывает, что обработка процедур, так построение пробит или логистической регрессионной модели для данных объемом 600 переменных и 3000 случаев занимает пару-тройку секунд.

² На данный момент справочная система доступна только на английском языке.

³ Работать с SDA возможно практически во всех современных браузерах, поддерживающих javascript. Брандмауэр пользователя не должен блокировать порт 8080. Некоторые браузеры интерпретируют дерево переменных как всплывающее окно и блокируют его появление. Поэтому при отсутствии дерева переменных, необходимо установить в настройках браузера возможность открытия всплывающих окон для используемого сайта SDA (<http://sophist.hse.ru:8080>).

опросов» раздела «Доступ к данным» выбрать соответствующую опцию (см. Рис. 1)¹.

Перед пользователем появится каталог исследований, опубликованных в SDA. В настоящий момент для анализа в этой системе доступно не очень большое количество опросов, но оно постоянно увеличивается. В представленном каталоге можно посмотреть описание исследования посредством перехода по соответствующей гиперссылке или приступить непосредственно к анализу данных, выбрав название нужного исследования, которое является его стартовой гиперссылкой. Для удобства пользователей мы привели непосредственно на страничке каталога краткое описание возможностей SDA.

2.2. Структура фреймов и общие принципы работы с ними

The screenshot shows the SDA web interface for a study titled "Условия работы ученых и привлекательность научной карьеры". The interface is divided into four main frames:

- Фрейм «Панель задач» (Task Panel):** Located at the top, it contains navigation tabs like "Analysis", "Create Variables", "Download", "Codebook", and "Getting Started". Below these are buttons for "Frequencies or crosstabulation", "Comparison of means", "Correlation matrix", "Comparison of correlations", "Multiple regression", "Logit/Probit regression", and "List values of individual cases".
- Фрейм «Выбор переменных» (Variable Selection):** Located on the left, it includes a "Variable Selection" section with a "Selected:" field (containing "Q1.2") and a "Copy to:" section with options for "Row", "Col", "Ctrl", and "Filter". It also has a "Mode:" section with "Append" and "Replace" options.
- Фрейм «Дерево переменных» (Variable Tree):** Located in the middle-left, it displays a hierarchical tree of variables under the heading "Условия работы ученых и привлекательность научной карьеры".
- Фрейм «Процедура» (Procedure):** Located on the right, it contains the "SDA Frequencies/Crosstabulation Program" settings. This includes fields for "REQUIRED Variable names to specify" (Row: Q1.2), "OPTIONAL Variable names to specify" (Column, Control, Selection Filter(s), Weight), and various options for "TABLE OPTIONS" (e.g., Percentaging, Confidence intervals, N of cases to display) and "CHART OPTIONS" (e.g., Type of chart, Bar chart options, Show percents, Palette, Size). It also has a "Title:" field and buttons for "Run the Table" and "Clear Fields".

Рисунок 2. Окно исследования, опубликованного SDA

После перехода по ссылке с названием исследования перед пользователем появляется основное окно программного комплекса (Рис. 2), разбитое на 4 фрейма², отображение границ которых зависит от настроек браузера:

- фрейм «Панель задач», расположенный вверху экрана и предназначенный для выбора процедур обработки и анализа данных, для вывода справочной информации о принципах работы в SDA, а также для перехода к некоторым разделам сайта Архива;
- фрейм «Выбор переменных», расположенный слева в верхней части и фрейм «Дерево переменных», расположенный под фреймом «Выбор

¹ Адрес сайта: <http://sophist.hse.ru/db/sda.shtml>

² Фрейм - отдельная, законченная интернет-страница, которая вместе с другими интернет-страницами может быть отображён в окне браузера.

переменных». Эти фреймы используются для ввода переменных в процедуры;

- фрейм «Процедура», расположенный справа, предназначенный для формирования и организации вывода результатов процедур, которые отображаются в отдельном окне или вкладке (в зависимости от настроек браузера пользователя).

Далее будут рассмотрены особенности работы с каждым из представленных фреймов¹.

2.2.1 Фрейм «Панель задач»

Фрейм «Панель задач» включает в себя панель навигации по сайту ЕАЭСД, расположенную в самом верху и позволяющую переходить к основным страницам сайта Архива, а ниже панель SDA, предназначенную для выбора процедур обработки и анализа данных. Пиктограмма-ссылка в панели навигации в правом верхнем углу позволяет возвращаться к каталогу опубликованных в SDA опросов. Далее, говоря о фрейме «Панель задач», мы будем иметь в виду именно панель SDA этого фрейма.

Фрейм «Панель задач» включает в себя 5 опций панели SDA, две из которых – «Анализ» (Analysis) и «Создание переменных» (Create Variables) имеют подразделы. Эти подразделы отображаются в виде выпадающего меню при щелчке левой клавишей мыши на соответствующей опции (см. Рис.3).

Опция «Загрузить» (Download) была перенесена нами в другой раздел сайта ЕАЭСД. Пользователь, желающий загрузить данные на свой компьютер, должен пройти регистрацию. После получения логина и пароля ему становится доступна возможность скачивать данные.

Опция «Коудбук» (Codebook) предназначена для просмотра списка одномерных линейных распределений, а опция «Приступая к работе» (Getting Started) позволяет получить раздел справки, посвященной общим принципам работы в SDA, которые отображаются в отдельном окне.

Analysis	Create Variables	Download	Codebook	Getting Started		
Frequencies or crosstabulation	Comparison of means	Correlation matrix	Comparison of correlations	Multiple regression	Logit/Probit regression	List values of individual cases

Рисунок 3. Подразделы опции «Анализ» фрейма «Панель задач»

В разделе «Анализ» сгруппированы 5 процедур статистического анализа (см. Рис.3): «Линейные распределения/кросстабуляция» (Frequencies/Crosstabulation), «Сравнение средних» (Comparison of Means) - сюда включены t-тест процедура и дисперсионный анализ, «Корреляционная матрица» (Correlation Matrix), «Сравнение корреляций» (Comparison of Correlations) – сюда помимо корреляционного анализа, включен анализ надежности, «Множественная регрессия» (Multiple Regression), «Логит/Пробит регрессия» (Logit\Probit Regression).

Analysis	Create Variables	Download	Codebook	Getting Started
	Recode variables	Compute a new variable	List/Delete Created Variables	

Рисунок 4. Подразделы опции «Создание переменных» фрейма «Панель задач»

¹ Название всех процедур и опций представлено в системе на английском языке.

В разделе «Создание переменных» представлены опции 2 процедур для подготовки данных «Перекодировать переменные» (Recode Variables) и «Вычислить переменную» (Compute a new variable), кроме того здесь расположена процедура «Просмотр/Удаление созданных переменных» (List/Delete variables) (см. Рис. 4).

2.2.2 Фреймы «Выбор переменных» и «Дерево переменных»

Фрейм «Дерево переменных» отображает структуру данных и всегда остается неизменным. Фрейм «Выбор переменных» (Variable Selection) представляет из себя инструмент ввода переменных во фрейм «Процедура». В зависимости от выбранной процедуры во фрейме «Панель задач», набор опций в нем будет отличаться (См. рис. 5).

Структура данных, представленная во фрейме «Дерево переменных» может включать в себя несколько уровней иерархии. Так, если одному вопросу в анкете соответствует несколько переменных, то такие переменные помещаются в раздел с названием вопроса. В качестве пиктограмм для разделов выступает открытая или закрытая книжки, указывающие на раскрытие или закрытие списков переменных. В качестве пиктограммы для отдельных переменных используется отдельный лист. После клика мышью на переменную, ее пиктограмма изменяется на лист отмеченный красной звездочкой, а имя отображается на панели «Выбрано» (Selected) фрейма «Выбор переменных».

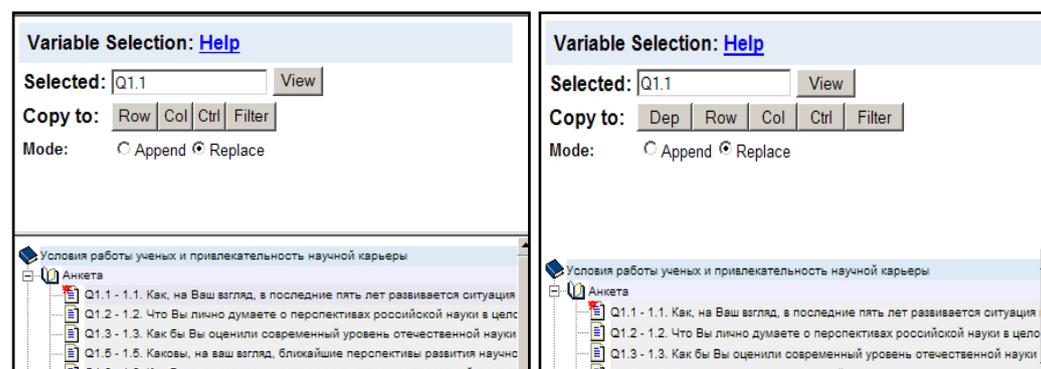


Рисунок 5. Фреймы «Выбор переменных» и «Дерево переменных» для процедур Линейные распределения/Кросстабуляция (слева) и «Сравнение средних» (справа)

Во фрейме «Выбор переменных» во всех случаях будет доступна лишь кнопка «Обзор» (View), позволяющая просматривать одномерные линейные распределения. Кнопки панели опций «Копировать» (Copy to) используется для копирования переменной в соответствующие панели ввода фрейма «Процедура». Переключатели опций панели «Режим» (Mode), позволяют указать возможность «Добавить» (Append) или «Заменить» (Replace) переменную в соответствующих полях фрейма «Процедура». SDA позволяет вводить имя переменной практически во все панели ввода вручную. К сожалению, попытка перетаскивания с помощью мыши переменной из дерева в какое-либо поле процедуры приводит не к копированию ее названия в это поле, а копированию ее названия с кодом javascript, что с нашей точки зрения, является недоработкой.

2.2.3 Фрейм «Процедура»

Фрейм «Процедура» может достаточно сильно различаться в зависимости от выбранной процедуры анализа или типа подготовки переменной. Для процедур анализа общим является то, что вверху этого фрейма располагаются панели ввода анализируемых переменных (см. Рис. 6).

The image shows two side-by-side software dialog boxes. The left box is titled "SDA Correlation Matrix Program" and contains fields for "Variables to Correlate" (1-16), "Missing-data exclusion" (Listwise/Pairwise), "Selection Filter(s)", "Weight", "Main statistic to display" (Pearson correlation), "Other statistics" (Alpha coefficient, Univariate stats, Std errors of correlations, Paired stats), "Index of Proportionality of Rows (PSQ)", "Reverse signs of correlations", "Color coding", and "Question text". It also has a "Title" field and "Run Correlations" and "Clear Fields" buttons. The right box is titled "SDA Frequencies/Crosstabulation Program" and contains fields for "REQUIRED Variable names to specify" (Row, Column), "OPTIONAL Variable names to specify" (Control, Selection Filter(s), Weight), "TABLE OPTIONS" (Percentaging: Column, Row, Total; Confidence intervals Level: 95 percent; Standard error of each percent; N of cases to display: Unweighted, Weighted; Summary statistics; Question text; Suppress table; Color coding; Show Z-statistic; Include missing-data values), "CHART OPTIONS" (Type of chart: Stacked Bar Chart; Orientation: Vertical, Horizontal; Visual Effects: 2-D, 3-D; Show percents; Palette: Color, Grayscale; Size: width, height), "Title", and "Run the Table" and "Clear Fields" buttons. Both boxes have a "Change number of decimal places to display" section at the bottom.

Рисунок 6. Фрейм «Процедура» для опций анализа «Линейные распределения/кросстабуляция» (слева), для «Корреляционная матрица» (справа)

Так, для таблиц сопряженности это будут переменные взятые по столбцу и строке, для регрессий зависимая переменная и предикторы и т.д. Чуть ниже, как правило, располагается панель «Выбор фильтра(ов)» (Selection Filter(s)), выбора переменной для разбиения на подвыборки - «Контрольная переменная» (Control) и «Взвешивание» (Weight). Под ними расположены опции для отображения и расчета параметров и коэффициентов, а справа - опции построения графиков, если такая возможность предусмотрена. Важно отметить, что для любой процедуры анализа в этом фрейме можно указать «Доверительный интервал» (Confidence intervals) в 90, 95 и 99 % (по умолчанию установлено значение 95%). Кроме того, всегда среди опций присутствует параметр «Цветовое кодирование» (Color Coding), предназначенный для выделения цветом значений и коэффициентов в зависимости от процедуры, с целью подчеркнуть различия между параметрами или отличие от ожидаемых значений. Еще ниже в этом фрейме расположена панель Title, предназначенная для ввода названия окна вывода процедуры, а также кнопка запуска процедуры, содержащая в своем названии слово «Run» и кнопка «очистки» полей «Clear fields». После запуска кнопки «Run» результат выполнения отображается в новом окне (или вкладке). В самом низу фрейма расположено меню выбора количества отображаемых десятичных знаков для результатов.

Confidence intervals

If this option is selected, an additional row of numbers is generated that contains the upper and lower bound of the confidence interval of the percentage (column, row, and/or total) in each cell. The confidence interval is the range of values within which the population value of the statistic is likely to fall. By default, the level of confidence is 95 percent, but the user can also select 99 percent or 90 percent.

The confidence interval is computed by converting the [standard error of each percentage](#) to a natural logarithm and then multiplying the log of the standard error by the value of Student's *t* appropriate to the level of confidence requested and to the number of [degrees of freedom](#). The result is added to the log of the percentage to obtain the upper bound of the confidence interval, and it is subtracted from the log of the percentage to obtain the lower bound. The logs of the upper bound and of the lower bound are then converted back to percentages (by taking the antilogs) and displayed in the table cell.

This conversion back and forth to logarithms results in confidence intervals that are asymmetric -- they are a little wider in the direction of 50% than in the direction of 0% or 100%. This is the same procedure used by Stata to calculate confidence intervals of percentages. Notice that the calculation of confidence intervals for a proportion (or for any mean) by the Comparison of Means program does not use this log transformation. Therefore, the confidence intervals calculated by the Comparison of Means program will be a little different than the confidence intervals calculated by the Crosstabulation program for the same proportions. This is also the case for Stata.

Alpha coefficient

Cronbach's alpha coefficient is a measure of how well the variables in the correlation matrix could be said to measure the same thing. If you added together all of the variables included in the correlation matrix to form a scale, alpha is the square of the correlation between the scale and the underlying factor.

The alpha coefficient is a function of the average correlation between the variables and of the number of variables. If some of the variables are scored in opposite directions, you should use the option to reverse the signs of some of the variables, so that a high score on all variables means the same thing.

Standard errors

A standard error for each correlation coefficient can be computed. If this option is requested, the standard errors are placed in a separate matrix, right under the matrix of correlation coefficients. **If the sample is more complex than a simple random sample, the standard errors calculated here are probably too small.**

The standard errors can be used to create confidence intervals for each correlation coefficient. For example, you can be 95% confident that the correlation coefficient in the population (for each pair of variables) is within the interval bounded by approximately two standard errors above and below the correlation coefficient calculated from the sample (as shown in the matrix). The actual multiple to use for creating confidence intervals is the *t*-statistic with (n-1) degrees of freedom.

The calculation of the standard error of the correlation coefficient in each cell is based by default on the UNWEIGHTED number of cases, even if a weight variable has been used for calculating the correlation coefficient. Ordinarily this

Рисунок 7. Разделы справочной системы с описанием опции «Доверительный интервал» (Confidence Interval) (слева) и «Коэффициент Альфа» (справа)

У пользователя всегда есть доступ к уже упомянутой нами качественной справочной системе - надо всего лишь кликнуть на гиперссылке с названием опции во фрейме «Процедура». На рис. 7 представлены окна вывода получения справки для опции «Доверительный интервал» процедуры «Линейные распределения/кросстабуляция» (в левой части) и «Коэффициент Альфа» (Alpha coefficient) для процедуры «Корреляционная матрица» (в правой части).

2.3. Подготовка данных для анализа

В SDA реализовано достаточно большое количество операций по подготовке переменных, носящих как временный, так и постоянный характер. Причем, есть возможность фильтровать и перекодировать переменные прямо в процедурах анализа. Как уже отмечалось выше, обычно для перекодирования и расчета переменных используется опция «Создание переменной» (Create Variable), расположенная во фрейме «Панель задач». Она содержит процедуры изменения данных двух типов: «Перекодировать переменные» (Recode variable) и «Вычислить переменную» (Compute variable) (см. раздел 2.2.1). Фильтр накладывается прямо во фрейме «Процедура» (см. раздел 2.2.3). При расчете переменных в SDA можно использовать:

- выражения условия IF / ELSE IF / ELSE, вложенные выражения условия и логические операторы EQ, NE, GT, GE, LT, LE, AND, OR;
- функции mean, sum, min, max, count, cum, missing;
- временные переменные \$temp;
- арифметические операторы + - * / ^ -var1 () ;
- арифметические функции ABS(x), EXP(x), LOG(x) or LN(x), LG10(x) or LOG10(x), MOD(x,a), RND(x) or ROUND(x), SQRT(x), TRUNC(x) ;
- функции случайного распределения;
- тригонометрические функции.

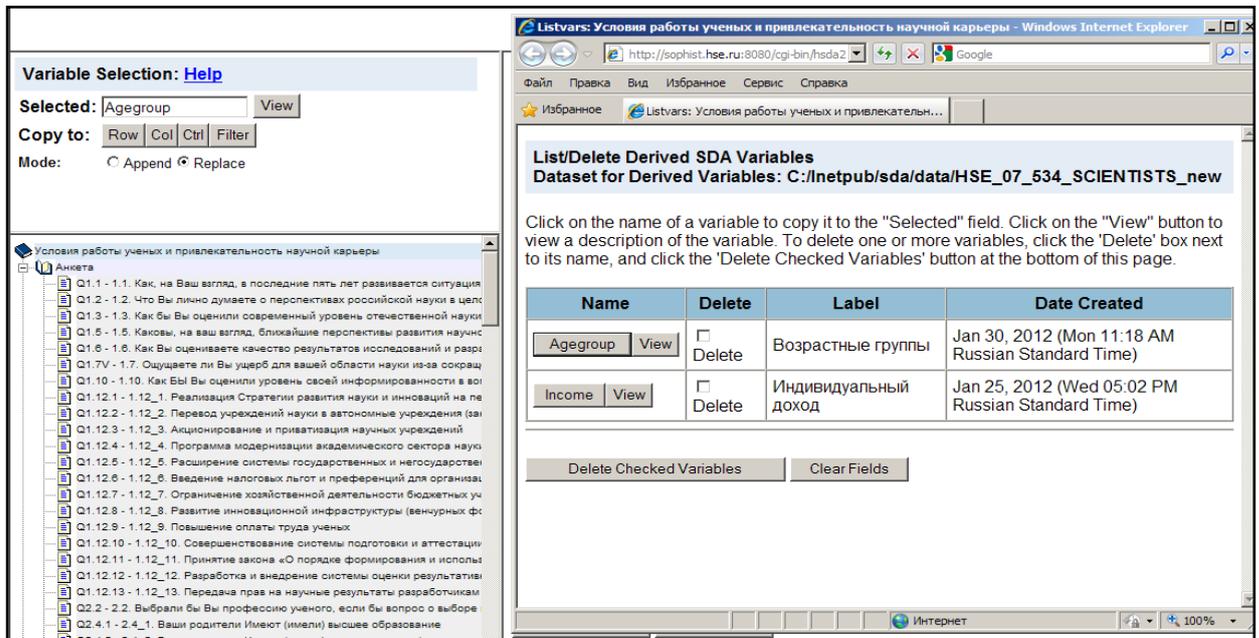


Рисунок 8 Фреймы «Выбор переменных» и «Дерево переменных (слева) и окно «Распечатать/удалить построенные SDA переменные» (справа)

К сожалению, в отличие от Nesstar, подготовленные переменные отображаются не в дереве переменных, а в отдельном окне переменных, подготовленных пользователем – «Распечатать/удалить построенные SDA переменные» (List/Delete Derived SDA Variables)¹, что, с нашей точки зрения, создает некоторые неудобства при работе. Для переноса переменной во фрейм «Выбор переменных» основного окна SDA необходимо нажать кнопку с именем переменной (на Рис. 8 это кнопка Agegroup).

2.4. Анализ данных

Для выбора процедур анализа используется опция «Анализ» (Analysis) из фрейма «Панель задач». Размеры данной статьи не позволяют нам осветить использование всех процедур анализа в SDA, поэтому мы ограничимся базовыми возможностями - теми, что с нашей точки зрения будут интересны пользователям в первую очередь.

2.4.1 Построение линейных распределений

Существует несколько способов просмотра линейных распределений в SDA, причем, отображаемые результаты будут различаться по форматам представления.

¹ Или вкладке в зависимости от настроек браузера пользователя.

<p>Title Page</p> <p>INDEXES</p> <p>Sequential Variable List</p> <p>Alphabetical Variable List</p>	<p>Q3.23 3.23. Как можно определить Ваше основное занятие в то время?</p> <table border="1"> <thead> <tr> <th>Percent</th> <th>N</th> <th>Value</th> <th>Label</th> </tr> </thead> <tbody> <tr> <td>3.1</td> <td>2</td> <td>1</td> <td>Работа в органах исполнительной и законодательной власти</td> </tr> <tr> <td>73.8</td> <td>48</td> <td>2</td> <td>Преподавательская работа в высшем учебном заведении</td> </tr> <tr> <td>1.5</td> <td>1</td> <td>3</td> <td>Работа в государственных организациях, кроме указанных в пп.</td> </tr> <tr> <td>9.2</td> <td>6</td> <td>4</td> <td>Работа в коммерческой структуре, кроме преподавательской, на</td> </tr> <tr> <td>1.5</td> <td>1</td> <td>5</td> <td>Самостоятельная предпринимательская деятельность</td> </tr> <tr> <td>3.1</td> <td>2</td> <td>6</td> <td>Обучение</td> </tr> <tr> <td>0.0</td> <td>0</td> <td>7</td> <td>Не работал(а)</td> </tr> <tr> <td>0.0</td> <td>0</td> <td>8</td> <td>Другое</td> </tr> <tr> <td>0.0</td> <td>0</td> <td>98</td> <td>Затрудняюсь ответить/отказ</td> </tr> <tr> <td>7.7</td> <td>5</td> <td>99</td> <td>Нет ответа</td> </tr> <tr> <td></td> <td>2,935</td> <td>.</td> <td>(No Data)</td> </tr> <tr> <td>100.0</td> <td>3,000</td> <td></td> <td>Total</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="2">Summary Statistics</th> </tr> </thead> <tbody> <tr> <td>Min = 1</td> <td>Mean = 9.800</td> </tr> <tr> <td>Max = 99</td> <td>Std. Dev. = 25.968</td> </tr> <tr> <td>Median = 2</td> <td>Variance = 674.350</td> </tr> </tbody> </table> <p>(Based on 65 valid cases)</p> <table border="1"> <thead> <tr> <th colspan="2">Properties</th> </tr> </thead> <tbody> <tr> <td>Data type:</td> <td>numeric</td> </tr> <tr> <td>Missing-data code:</td> <td>F2</td> </tr> <tr> <td>Record/columns:</td> <td>1/4602-4603</td> </tr> </tbody> </table>	Percent	N	Value	Label	3.1	2	1	Работа в органах исполнительной и законодательной власти	73.8	48	2	Преподавательская работа в высшем учебном заведении	1.5	1	3	Работа в государственных организациях, кроме указанных в пп.	9.2	6	4	Работа в коммерческой структуре, кроме преподавательской, на	1.5	1	5	Самостоятельная предпринимательская деятельность	3.1	2	6	Обучение	0.0	0	7	Не работал(а)	0.0	0	8	Другое	0.0	0	98	Затрудняюсь ответить/отказ	7.7	5	99	Нет ответа		2,935	.	(No Data)	100.0	3,000		Total	Summary Statistics		Min = 1	Mean = 9.800	Max = 99	Std. Dev. = 25.968	Median = 2	Variance = 674.350	Properties		Data type:	numeric	Missing-data code:	F2	Record/columns:	1/4602-4603
Percent	N	Value	Label																																																																		
3.1	2	1	Работа в органах исполнительной и законодательной власти																																																																		
73.8	48	2	Преподавательская работа в высшем учебном заведении																																																																		
1.5	1	3	Работа в государственных организациях, кроме указанных в пп.																																																																		
9.2	6	4	Работа в коммерческой структуре, кроме преподавательской, на																																																																		
1.5	1	5	Самостоятельная предпринимательская деятельность																																																																		
3.1	2	6	Обучение																																																																		
0.0	0	7	Не работал(а)																																																																		
0.0	0	8	Другое																																																																		
0.0	0	98	Затрудняюсь ответить/отказ																																																																		
7.7	5	99	Нет ответа																																																																		
	2,935	.	(No Data)																																																																		
100.0	3,000		Total																																																																		
Summary Statistics																																																																					
Min = 1	Mean = 9.800																																																																				
Max = 99	Std. Dev. = 25.968																																																																				
Median = 2	Variance = 674.350																																																																				
Properties																																																																					
Data type:	numeric																																																																				
Missing-data code:	F2																																																																				
Record/columns:	1/4602-4603																																																																				

Рисунок 9. Окно коудбука

Первой возможностью, является просмотр линейных распределений в коудбуке, для открытия которого необходимо выбрать соответствующую опцию “Коудбук” во фрейме «Панель задач» (выбор данной опции описан в разделе 2.2.1 статьи). Откроется новое окно (Рис.9), в котором переменные представлены либо по алфавиту (Alphabetical Variable List), либо в порядке ввода в массив данных (Sequential Variable List). При этом сгруппированные в разделы переменные, например вопросы со множественными ответами, будут представлены в конце списка. Уже просмотренные линейные распределения помечаются как посещенные гиперссылки. Линейные распределения отображаются здесь вместе с суммарными статистиками (мерами центральной тенденции, стандартным отклонением и дисперсией) и свойствами переменной. В коудбуке последовательно можно просмотреть линейные распределения почти всех переменных. Однако, стоит отметить, что для переменных с большим количеством значений (таких как возраст), в коудбуке отображаются только диапазон значений переменной и суммарные статистики.

Вторым способом просмотра линейных распределений является выбор во фрейме «Дерево переменных» в основном окне SDA имени переменной и последующее нажатие во фрейме «Выбор переменных» на кнопку «Просмотр» (о чем мы уже упоминали в разделе 2.2.2). В этом случае отобразятся линейные распределения со свойствами переменных, но без суммарных статистик. Для переменных с большим количеством значений (таких как возраст), здесь в отличие от коудбука будут представлены линейные распределения (Рис. 10).

Percent	N	Value	Label
6.6	199	1	Произойдет общее улучшение ситуации
30.6	917	2	Произойдет улучшение ситуации в отдельных областях науки
28.6	858	3	Произойдет улучшение ситуации в отдельных научных организаци
19.9	596	4	Ничего не изменится
11.2	337	5	Произойдет общее ухудшение ситуации
2.9	87	6	Наиболее вероятен ее полный распад
0.2	6	9	Нет ответа
100.0	3,000		Total

Properties	
Data type:	numeric
Missing-data codes:	F1
Mean:	3.08
Std Dev:	1.24
Record/column:	1/172

Selected Study: Условия работы ученых и привлекательность научной карьеры

Рисунок 10. Линейные распределения выводимые поле нажатия на кнопку «Просмотр» во фрейме «Выбор переменных»

Одномерное линейное распределение					
SDA 3.5: Tables					
Условия работы ученых и привлекательность научной карьеры					
Feb 17, 2012 (Fri 08:42 PM Russian Standard Time)					
Variables					
Role	Name	Label	Range	MD	Dataset
Row	Q1.2	1.2. Что Вы лично думаете о перспективах российской науки в целом на ближайшие пять лет?	1-9	F1	1
Frequency Distribution					
Cells contain:					Distribution
-Column percent					
-N of cases					
Q1.2	1: Произойдет общее улучшение ситуации				6.6 199
	2: Произойдет улучшение ситуации в отдельных областях науки				30.6 917
	3: Произойдет улучшение ситуации в отдельных научных организаци				28.6 858
	4: Ничего не изменится				19.9 596
	5: Произойдет общее ухудшение ситуации				11.2 337
	6: Наиболее вероятен ее полный распад				2.9 87
	9: Нет ответа				.2 6
	COL TOTAL				100.0 3,000
Summary Statistics					
Mean =	3.08	Std Dev =	1.24	Coef var =	.40
Median =	3.00	Variance =	1.55	Min =	1.00
Mode =	2.00	Skewness =	.57	Max =	9.00
Sum =	9,252.00	Kurtosis =	.30	Range =	8.00
<i>Inference about the mean:</i>					
Std Err =	.02	CV(mean) =	.01		
Allocation of cases					
Valid cases	3,000				
Total cases	3,000				

Рисунок 11. Линейные распределения выводимые в результате запуска процедуры «Линейные распределения/Кросстабуляция»

Наконец, третьей (и наиболее полной) возможностью является построение линейных распределений с помощью процедуры «Линейные распределения/Кросстабуляция», настройки которой по умолчанию отображаются в правом фрейме при открытии стартовой ссылки массива. Если в настройке процедуры указывается только переменная «По строке» («Row»), и не указывается переменная «По столбцу» («Column») – то после нажатия на кнопку «Построить таблицу» (Run the table), выполняется построение линейных распределений. Следует отметить, что переменная, указываемая в поле «По строке», в рамках данной процедуры всегда считается зависимой и, если ее не указать, результат выводится не будет. Процедура «Линейные распределения/Кросстабуляция» допускает настройку параметров: есть возможность отображать линейные распределения по подвыборкам путем определения контрольной переменной и/или фильтра, строить графики, рассчитывать различные статистики. Получаемый результат представлен на Рис. 11.

2.4.2 Построение таблиц сопряженности

Построение таблиц осуществляется с помощью той же процедуры, что и линейных распределений, и отличается только тем, что в поле «По колонке» (Column) вводится дополнительная переменная (верхняя часть Рис. 12). Процедура предоставляет все стандартные возможности, а именно:

- возможность вывода наблюдаемого числа случаев для каждой клетки таблицы (N of cases to display) как взвешенного, так и не взвешенного¹;
- расчет процента по строке (Row),
- расчет процента по столбцу (Column)
- расчет табличного процента (Total),
- расчет стандартной ошибки среднего (Standart errors of each percent),
- расчет доверительного интервала.

Для вычисления коэффициентов парной связи нужно выбрать опцию «Суммарные статистики» (Summary statistics). Рассчитываются такие коэффициенты, как критерий хи-квадрат Пирсона (Chisq-P), коэффициенты Гамма (Gamma), тау b и с Кендалла (Tau-b, Tau-c), d Соммера (Sommers' d), коэффициент корреляции r Пирсона (R) и коэффициент нелинейных отношений (Eta). Кроме того, реализована возможность отображения стандартизованных остатков для ячеек таблицы: «Показать Z-коэффициенты» (Show Z-statistic).

Пользователю предоставлена возможность управлять параметрами отображения таблицы, например «Отключить отображение таблицы» (Suppress Table).

¹ Для отображения взвешенных значений разумеется в соответствующем поле «Взвешивание» (Weight) должен быть указан весовой коэффициент.

 Главная Карта сайта
 
 SDA

SDA [Use classic interface] Selected Study: **Условия работы ученых и привлекательность научной карьеры**

Analysis Create Variables Download Codebook Getting Started

Variable Selection: [Help](#)
 Selected: Q1.1 View
 Copy to: Row Col Ctrl Filter
 Mode: Append Replace

SDA Frequencies/Crosstabulation Program
 Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify
 Row: Q1.2(*-6)
OPTIONAL Variable names to specify
 Column: Q1.1(*-6)
 Control:
 Selection Filter(s): Example: age(18-50)
 Weight:

TABLE OPTIONS
Percentaging:
 Column Row Total
 Confidence intervals Level: 95 percent
 Standard error of each percent
N of cases to display:
 Unweighted Weighted
 Summary statistics
 Question text Suppress table
 Color coding Show Z-statistic
 Include missing-data values

CHART OPTIONS
Type of chart: (No Chart)
Bar chart options:
 Orientation: Vertical Horizontal
 Visual Effects: 2-D 3-D
Show percents: Yes
Palette: Color Grayscale
Size - width: 600 height: 400

Title: Таблица сопряженности
 Run the Table Clear Fields

Условия работы ученых и привлекательность научной карьеры
 Анкета
 Q1.1 - 1.1. Как, на Ваш взгляд, в последние пять лет развивается ситуация
 Q1.2 - 1.2. Что Вы лично думаете о перспективах российской науки в целом на ближайшие пять лет?
 Q1.3 - 1.3. Как бы Вы оценили современный уровень отечественной науки
 Q1.5 - 1.5. Каковы, на ваш взгляд, ближайшие перспективы развития научных исследований в России?
 Q1.6 - 1.6. Как Вы оцениваете качество результатов исследований и разработок в области науки?
 Q1.7V - 1.7. Ощущаете ли Вы ущерб для вашей области науки из-за сокращения финансирования?
 Q1.10 - 1.10. Как бы Вы оценили уровень своей информированности в вопросах развития науки и инноваций?
 Q1.12.1 - 1.12_1. Реализация Стратегии развития науки и инноваций на период 2013-2020 гг.
 Q1.12.2 - 1.12_2. Перевод учреждений науки в автономные учреждения (в том числе в форме государственных учреждений)
 Q1.12.3 - 1.12_3. Акционирование и приватизация научных учреждений
 Q1.12.4 - 1.12_4. Программа модернизации академического сектора науки
 Q1.12.5 - 1.12_5. Расширение системы государственных и негосударственных научных организаций
 Q1.12.6 - 1.12_6. Введение налоговых льгот и преференций для организаций
 Q1.12.7 - 1.12_7. Ограничение хозяйственной деятельности бюджетных учреждений
 Q1.12.8 - 1.12_8. Развитие инновационной инфраструктуры (венчурных фондов)
 Q1.12.9 - 1.12_9. Повышение оплаты труда ученых
 Q1.12.10 - 1.12_10. Совершенствование системы подготовки и аттестации научных кадров
 Q1.12.11 - 1.12_11. Принятие закона «О порядке формирования и использования государственного задания»
 Q1.12.12 - 1.12_12. Разработка и внедрение системы оценки результатов научных исследований и разработок
 Q1.12.13 - 1.12_13. Передача прав на научные результаты разработчикам
 Q2.2 - 2.2. Выбрали бы Вы профессию ученого, если бы вопрос о выборе профессии зависел только от Вас?
 Q2.4.1 - 2.4_1. Ваши родители Имеют (имели) высшее образование
 Q2.4.2 - 2.4_2. Ваши родители Имеют (имели) ученую степень (кандидат и доктор наук)

Variables						
Role	Name	Label	Range	MD	Dataset	
Row	Q1.2(*-6)	1.2. Что Вы лично думаете о перспективах российской науки в целом на ближайшие пять лет?	1-9	F1	1	
Column	Q1.1(*-6)	1.1. Как, на Ваш взгляд, в последние пять лет развивается ситуация в науке	1-9	F2	1	

Frequency Distribution								
		Q1.1						
		1	2	3	4	5	6	ROW TOTAL
Cells contain:		Наблюдается значительное улучшение ситуации	Наблюдается некоторое улучшение ситуации	Сохраняется стабильная ситуация	Сохраняется кризисная ситуация	Наблюдается некоторое ухудшение ситуации	Наблюдается значительное ухудшение ситуации	
-Row percent								
-N of cases								
Q1.2	1: Произойдет общее улучшение ситуации	20.6 41	58.3 116	9.0 18	10.1 20	1.5 3	.5 1	100.0 199
	2: Произойдет улучшение ситуации в отдельных областях науки	5.0 46	52.7 483	12.2 112	25.5 234	3.1 28	1.5 14	100.0 917
	3: Произойдет улучшение ситуации в отдельных научных организациях	1.2 10	32.2 276	20.3 174	40.3 346	5.1 44	.9 8	100.0 858
	4: Ничего не изменится	.2 1	9.2 55	22.3 133	53.5 319	11.2 67	3.5 21	100.0 596
	5: Произойдет общее ухудшение ситуации	.6 2	4.7 16	2.4 8	54.9 185	16.3 55	21.1 71	100.0 337
	6: Наиболее вероятен ее полный распад	1.1 1	2.3 2	.0 0	33.3 29	4.6 4	58.6 51	100.0 87
COL TOTAL		3.4 101	31.7 948	14.9 445	37.8 1,133	6.7 201	5.5 166	100.0 2,994
Means		1.81	2.34	3.00	3.44	3.77	4.81	3.07
Std Devs		.92	.84	.89	1.12	1.10	1.21	1.22

Summary Statistics
 Eta* = .57 Gamma = .61 Chisq-P(25) = 1,586.59 (p= 0.00)
 R = .56 Tau-b = .47 Chisq-LR(25) = 1,295.30 (p= 0.00)
 Somers' d* = .48 Tau-c = .42
 *Row variable treated as the dependent variable.

Allocation of cases
 Valid cases 2,994
 Cases with invalid codes on row or column variable 6
 Total cases 3,000

Рисунок 12. Часть основного окна SDA (вверху) и часть окна вывода (внизу) для процедуры «Линейные распределения/Кросстабуляция»

2.4.3 Регрессионное моделирование

В качестве примера более сложных процедур статистического анализа, которые можно выполнять в SDA, приведем пример построения множественной и логистической регрессии. Напомним, что выбор этих процедур анализа осуществляется во фрейме «Панель задач» (подробнее см. раздел 2.2.1), а ввод во фрейм переменных посредством фреймов «Выбор переменных» и «Дерево переменных» (раздел 2.2.2 статьи).

Multiple regression	Logit/Probit regression	List values of individual cases			
SDA Multiple Regression Program Help: General / Dummy vars / Product terms Dependent: Q3.52.4 (0-100) Independent: (You can tab from one input box to the next) 1: Q3.52.1 (0-100) 2: Q3.52.2 (0-100) 3: Q3.52.3 (0-100) 4: Q3.52.5 (0-100) 5: 6: Q3.52.6 (0-100) 7: Q3.52.7 (0-100) 8: Q3.52.8 (0-100) 9: Q3.52.9 (0-100) 10: 11: 12: 13: 14: 15: 16: More independent variables Selection Filter(s): TYPE (1) Example: age(18-50) Weight: Other statistics: <input checked="" type="checkbox"/> T-tests <input checked="" type="checkbox"/> Global tests <input type="checkbox"/> Confidence intervals Level: 95 percent <input type="checkbox"/> Univariate stats <input type="checkbox"/> Product: B * Mean Matrices to display: <input checked="" type="checkbox"/> Correlation matrix <input type="checkbox"/> Covariance matrix <input type="checkbox"/> Covariance matrix of coefficients Other options: <input type="checkbox"/> Question text <input type="checkbox"/> Color coding <input type="checkbox"/> Suppress independent variables list Title: Множественная регрессия Run Regression Clear Fields Change number of decimal places to display: For coefficients: 3 For t-tests: 3 For F-test: 3 For univariate stats and B*Mean: 3 For correlation matrix: 2 For covariance matrix: 2 For covariance matrix of coefficients: 6 More independent variables 17: 18: 19: 20:			SDA Logit/Probit Regression Program Help: General / Dummy vars / Product terms Type of Regression: <input checked="" type="radio"/> Logit <input type="radio"/> Probit Dependent: Q1.2 (d.1-2) Independent: (You can tab from one input box to the next) 1: Q1.6 (d.1-2) 2: Q5.2 (21-60) 3: 4: 5: 6: 7: 8: 9: 10: 11: 12: 13: 14: 15: 16: More independent variables Selection Filter(s): TYPE (1) Example: age(18-50) Weight: Other statistics: <input checked="" type="checkbox"/> T-tests <input checked="" type="checkbox"/> Global tests <input type="checkbox"/> Confidence intervals Level: 95 percent <input checked="" type="checkbox"/> Exp(B) - for logit <input type="checkbox"/> Univariate stats <input type="checkbox"/> Product: B * Mean <input type="checkbox"/> Covariance matrix of coefficients Other options: <input type="checkbox"/> Question text <input type="checkbox"/> Color coding <input type="checkbox"/> Suppress independent variables list Title: Run Logit/Probit Clear Fields Change number of decimal places to display: For coefficients: 3 For t-tests: 3 For F-test: 3 For univariate stats and B*Mean: 3 For covariance matrix of coefficients: 6 More independent variables 17: 18: 19: 20: 21: 22: 23: 24: 25: 26: 27: 28: 29: 30: 31: 32:		

Рисунок 13. Часть фрейма «Процедура» для процедур анализа «Множественная регрессия» (слева), и «Логит\Пробит регрессия» (справа)

Фреймы «Процедура» для построения регрессионных моделей не сильно отличаются друг от друга (Рис. 13). Для построения необходимой модели нужно выбрать соответствующие опции в меню «Тип регрессии» (Type regression) в верхней части фрейма. Вверху также расположены панель для ввода зависимой переменной (Dependent), а ниже - панели для ввода независимых переменных (Independent). Количество независимых переменных, которые можно вводить в регрессионные уравнения достаточно велико: пользователю предоставляется 16 основных панелей ввода независимых переменных сверху фрейма. Еще 35 дополнительных панелей находятся внизу фрейма в разделе «Дополнительные независимые переменные» (More independent variables). Для перехода в эту панель можно использовать гиперссылку «More independent variables» вверху фрейма.

Панели переменных допускают временную перекодировку, а именно:

- в скобках после имени переменной можно указать диапазон значений, которые будут учитываться при анализе;
- допускается возможность временной перекодировки исходных переменных в дихотомические. Для этого необходимо после имени переменной и символа «d:» в скобках указывать диапазон значений, которые перекодируются в 1, остальные значения будут перекодированы в 0;
- если при построении логистической или пробит модели в качестве зависимой указать не дихотомическую переменную, то при выполнении расчетов такая переменная автоматически перекодируется в бинарную. При этом, наименьшее значение исходной переменной перекодируются в 0, а остальные в 1.

Для всех видов регрессий рассчитываются стандартные коэффициенты. Для множественной регрессии это коэффициенты B и стандартная ошибка B ($SE(B)$), стандартизованные коэффициенты бета ($Beta$) и стандартная ошибка коэффициента бета ($SE(Beta)$), коэффициент корреляции R Пирсона ($Multiple R$), коэффициенты детерминации R -квадрат без поправок (R -Squared) и с поправками ($Adjusted R$ -Square), стандартная ошибка оценки (SE of Estimate ($Root$ MSE)) (см. Рис. 14).

Для логистической и пробит-регрессии рассчитываются коэффициенты B , стандартная ошибка B ($SE(B)$), а также коэффициент логарифмического правдоподобия (Log Likelihood) и псевдо R -квадрат Кокса и Снелла ($Pseudo R$ -sq) (см. Рис. 15).

Также можно, осуществив выбор требуемых параметров в меню «Другие статистики» ($Other$ statistics) и «Отображение матриц» ($Matrix$ to display) рассчитать дополнительные статистики для всех видов регрессий:

- «Т-тест» для средних (T -test на Рис. 10 и 11);
- Статистики Вальда с помощью опции «Общий тест» ($Global$ test);
- «Одномерный критерий» ($Univariate$ test);
- «Произведение коэффициента B на среднее значение переменной» ($Product: B * Mean$).
- «Матрицу ковариат для коэффициентов» ($Covariance$ matrix of coefficients)
Кроме того можно задать:
- вывод корреляционной и ковариационной матрицы с помощью соответствующих опций меню «Отображение матриц» фрейма «Процедура» для множественной регрессии;
- Вывод антилога для логистической регрессии, посредством выбора опции «Антилог» ($Exp(B)$ - for logit) во фрейме «Процедура».

Множественная регрессия

SDA 3.5: Regression

Условия работы ученых и привлекательность научной карьеры

Feb 20, 2012 (Mon 07:40 PM Russian Standard Time)

Variables						
Role	Name	Label	Range	MD	Dataset	
Dependent	Q3.52.4(0-100)	3.52_4. Сколько рабочего времени Вы тратите на Преподавание, научное руководство молодыми специалистами, аспирантами, соискателями и т.д	0-999	F3	1	
Independent	Q3.52.1(0-100)	3.52_1. Сколько рабочего времени Вы тратите на Фундаментальные исследования	0-999	F3	1	
Independent	Q3.52.2(0-100)	3.52_2. Сколько рабочего времени Вы тратите на Прикладные исследования и разработки	0-999	F3	1	
Independent	Q3.52.3(0-100)	3.52_3. Сколько рабочего времени Вы тратите на Выполнение административных функций	0-999	F3	1	
Independent	Q3.52.5(0-100)	3.52_5. Сколько рабочего времени Вы тратите на Редактирование научных материалов, отчетов, публикаций	0-999	F3	1	
Independent	Q3.52.6(0-100)	3.52_6. Сколько рабочего времени Вы тратите на Выполнение вспомогательных функций	0-999	F3	1	
Independent	Q3.52.7(0-100)	3.52_7. Сколько рабочего времени Вы тратите на Участие в совещаниях, семинарах и т.д. по вопросам научно-исследовательской деятельности	0-999	F3	1	
Independent	Q3.52.8(0-100)	3.52_8. Сколько рабочего времени Вы тратите на Участие в иных совещаниях, семинарах и т.д.	0-999	F3	1	
Independent	Q3.52.9(0-100)	3.52_9. Сколько рабочего времени Вы тратите на Другое	0-999	F3	1	
Filter	TYPE(1)	Тип организации(=Научно-исследовательский институт (центр))	1-4	F1	1	

Regression Coefficients				Test That Each Coefficient = 0		
	B	SE(B)	Beta	SE (Beta)	T-statistic	Probability
Q3.52.1(0-100)	-1.000	.000	-3.932	.000	-336397126.	.000
Q3.52.2(0-100)	-1.000	.000	-3.572	.000	-336679626.	.000
Q3.52.3(0-100)	-1.000	.000	-2.184	.000	-309950849.	.000
Q3.52.5(0-100)	-1.000	.000	-1.438	.000	-279519005.	.000
Q3.52.6(0-100)	-1.000	.000	-1.733	.000	-305711937.	.000
Q3.52.7(0-100)	-1.000	.000	-.942	.000	-219739274.	.000
Q3.52.8(0-100)	-1.000	.000	-.715	.000	-194157149.	.000
Q3.52.9(0-100)	-1.000	.000	-.659	.000	-189286072.	.000
Constant	100.000	.000			351.311,574.714	.000

Multiple R = 1.000 R-Squared = 1.000 Adjusted R-Squared = 1.000 SE of Estimate (Root MSE) = .000

Allocation of cases

Valid cases 1,533

Cases excluded by filter 1,401

Cases with invalid codes on variables in the analysis 66

Total cases 3,000

Missing data excluded: Listwise

Рисунок 14. Часть окна вывода процедуры анализа «Множественная регрессия»

SDA 3.5: Logit/Probit Regression

Условия работы ученых и привлекательность научной карьеры

Feb 20, 2012 (Mon 07:48 PM Russian Standard Time)

Variables						
Role	Name	Label	Range	MD	Dataset	
Dependent	Q1.2(d:1-2)	1.2. Что Вы лично думаете о перспективах российской науки в целом на ближайшие пять лет?	0-1		1	
Independent	Q1.6(d:1-2)	1.6. Как Вы оцениваете качество результатов исследований и разработок, выполненных за последние три года Вашей организацией?	0-1		1	
Independent	Q5.2(21-60)	5.2. Сколько полных лет Вам исполнилось?	21-99	F2	1	

Logit Coefficients			Test That Each Coefficient = 0	
	B	SE(B)	T-statistic	Probability
Q1.6(d:1-2)	.970	.089	10.919	.000
Q5.2(21-60)	-.017	.004	-4.743	.000
Constant	-.187	.161	-1.167	.244

Log Likelihood = -1,456.234 Pseudo R-sq = .051

Recode for 'Q1.2'
1 = 1-2; 0 = *.*

Recode for 'Q1.6'
1 = 1-2; 0 = *.*

Allocation of cases

Valid cases 2,289

Cases with invalid codes on variables in the analysis 711

Total cases 3,000

Рисунок 15. Часть окна вывода процедуры анализа «Логит\Пробит регрессия»

3. Заключение

Рамки статьи не позволяют описать такой сложный программный комплекс как SDA в деталях. Наша задача заключалась в том, чтобы привлечь внимание исследовательского сообщества к этому программному продукту и к новому сервису, впервые реализованному в России: ЕАЭСД – первый российский пользователь SDA.

Этот программный комплекс обладает рядом особенностей, на которые мы хотим еще раз обратить внимание.

1. Важнейшее преимущество данного программного комплекса (в сравнении, скажем, с SPSS) заключается в том, что пакет устанавливается один раз на сервере архива данных. Пользователи архива получают возможность пользоваться как самим пакетом, так и данными, не скачивая их на свой компьютер, что особенно важно для российских регионов.
2. Программный пакет обладает набором полезных пользовательских свойств, делающих его использование простым и удобным - «дружественный» интерфейс, удобная справочная система и т.д.;
3. Набор реализованных статистических процедур, хотя и не полон, но достаточен для проведения необходимого статистического анализа и решения основных исследовательских задач;
4. Пакет работает быстро, что позволяет анализировать большие массивы данных в режиме удаленного доступа, больше внимания уделяя интерпретации полученных результатов.

SDA -динамично развивающийся продукт, спектр его возможностей будет, согласно планам разработчиков, постоянно расширяться. Именно поэтому мы выбрали этот пакет для реализации одного из сервисов ЕАЭСД – обеспечения статистического анализа в режиме удаленного доступа. Мы уверены, что этот инновационный сервис, который ЕАЭСД предлагает научному сообществу, завоеует свою аудиторию в России, как он уже сделал это на Западе.



ДИЗАЙН, ВЕРСТКА
К. П. ЛАЗЕБНАЯ

