ОБЗОР ВОЗМОЖНОСТЕЙ SDA.

Митюшин А.А.

1. ВВЕДЕНИЕ

Характерными особенностями системы SDA являются:

- 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 процедуры 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).

Программный комплекс достаточно прост в освоении, так как обладает интуитивно понятным интерфейсом. Не последнюю роль в этом играет качественная, хорошо продуманная контекстная справочная система, использование которой помогает с легкостью ориентироваться в функциях и процедурах программного комплекса.

2. ИСПОЛЬЗОВАНИЕ SDA

2.1 НАЧАЛО РАБОТЫ

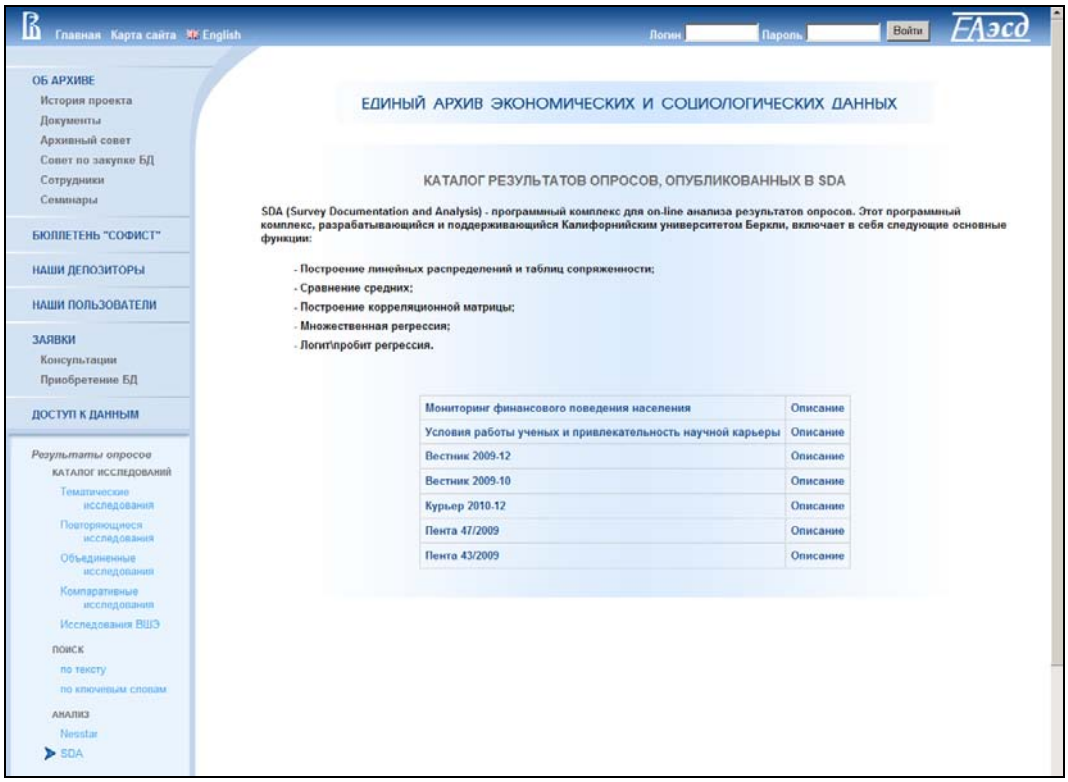


Рис. 1. Каталог исследований опубликованный в SDA

Для того, чтобы войти систему SDA необходимо в левой панели меню сайта ЕАЭСД в подразделе «Результаты опросов» раздела «Доступ к данным» выбрать соответствующую опцию (см. Рис. 1) или ввести в адресной строке браузера http://sophist.hse.ru/db/sda.shtml.

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 гипперссылке или приступить непосредственно к анализу данных, выбрав название нужного исследования, которое является его стартовой гипперссылкой. Для удобства пользователей на страничке каталога приведено краткое описание возможностей SDA.

2.2 СТРУКТУРА ФРЕЙМОВ И ОБЩИЕ ПРИНЦИПЫ РАБОТЫ С НИМИ

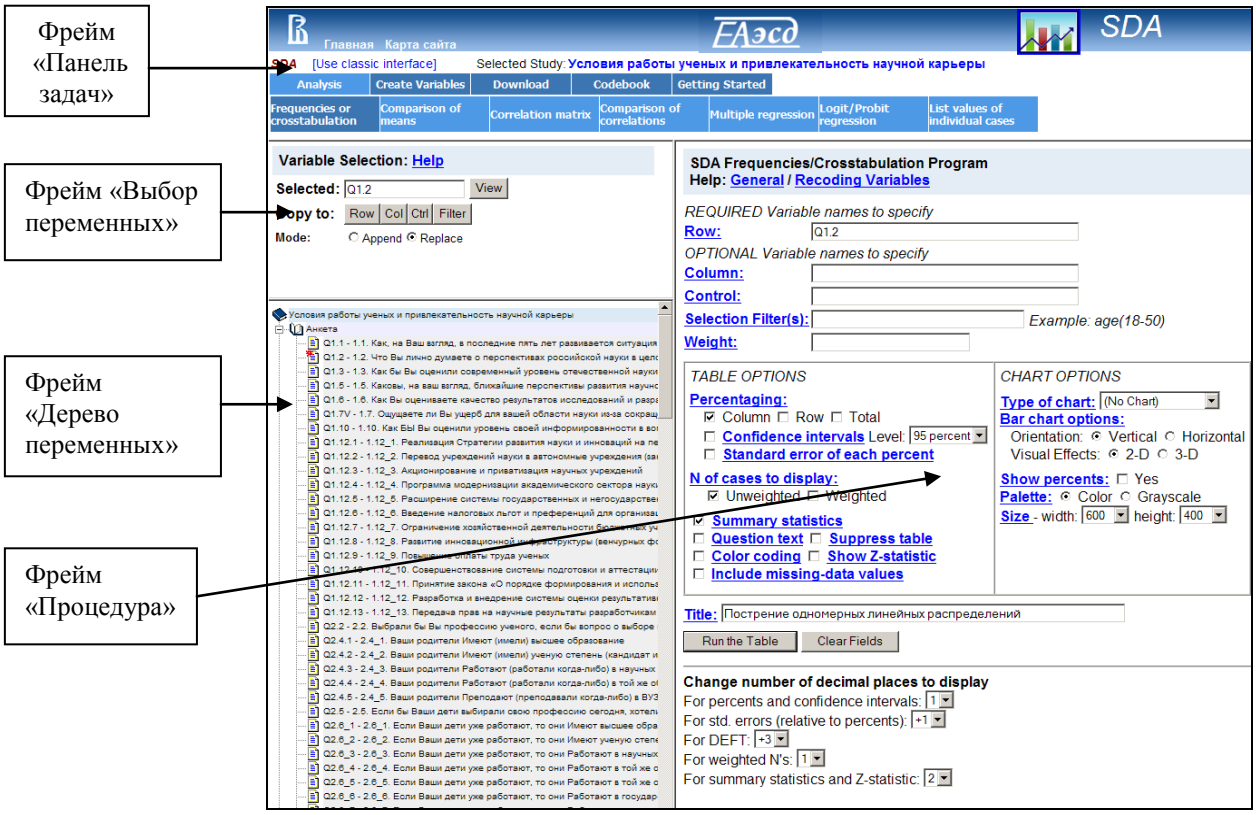


Рис. 2. Окно исследования, опубликованного SDA

После перехода по ссылке с названием исследования перед пользователем появляется основное окно программного комплекса (Рис. 2), разбитое на 4 фрейма, 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 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).

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.

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.

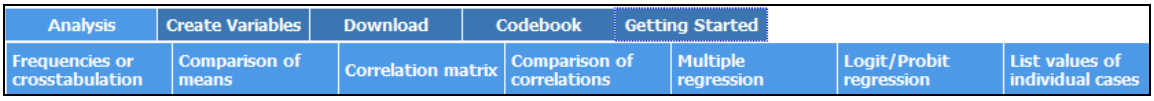


Рис. 3 Подразделы опции «Анализ» фрейма «Панель задач»

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).

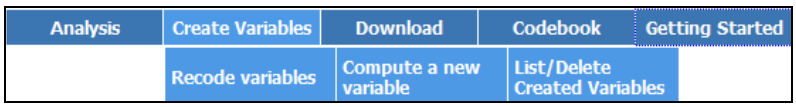


Рис. 4 Подразделы опции «Создание переменных» фрейма «Панель задач»

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 щелчка левой кнопки мыши по переменной, ее пиктограмма изменяется на лист отмеченный красной звездочкой, а имя отображается на панели «Выбрано» (Selected) фрейма «Выбор переменных».

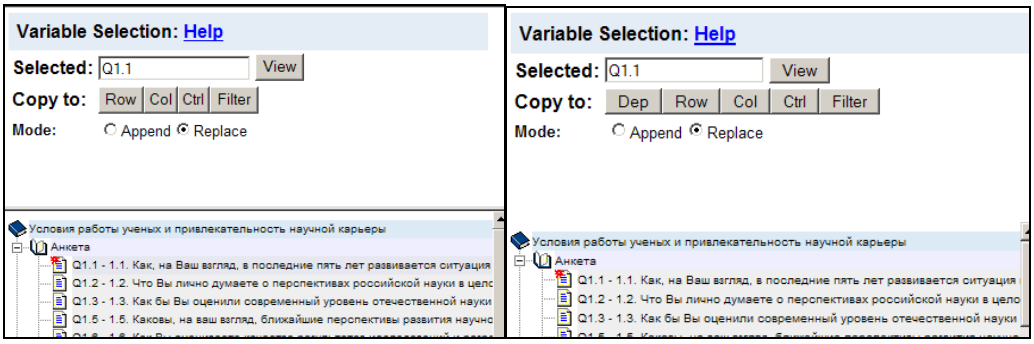


Рис. 5. Фреймы «Выбор переменных» и «Дерево переменных» для процедур Линейные распределения/Кросстабуляция (*слева*) и «Сравнение средних» (*справа*)

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. 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.

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).

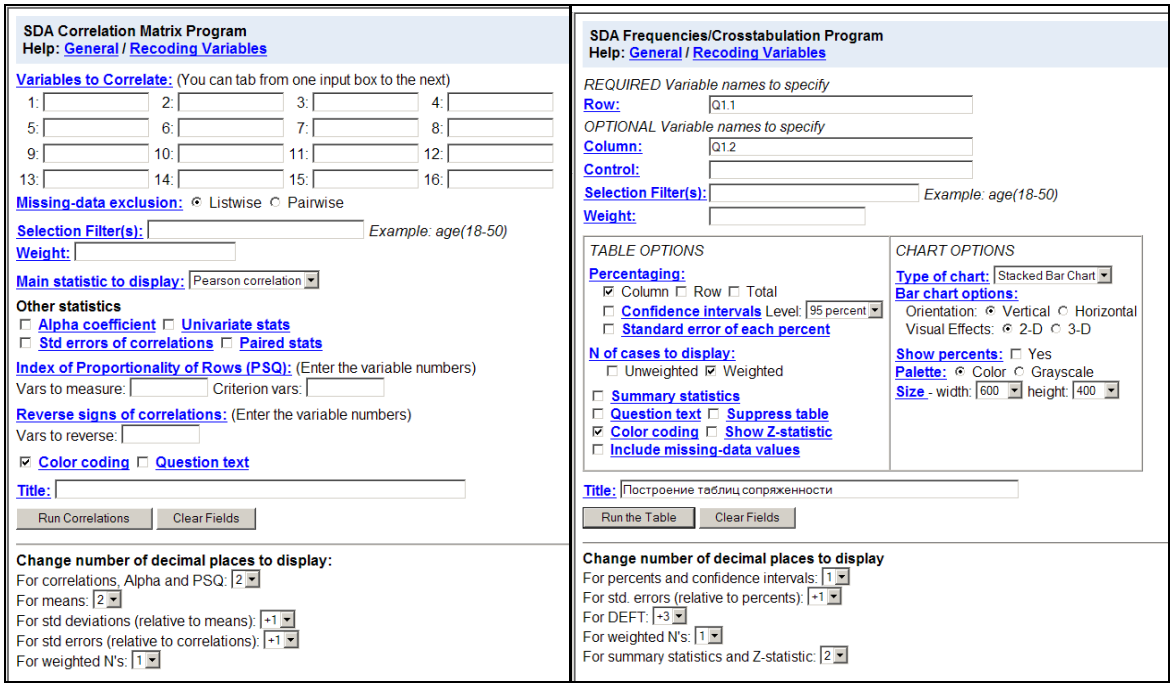


Рис. 6. Фрейм «Процедура» для опций анализа «Линейные распределения/кросстабуляция» (*слева*), для «Корреляционная матрица» (*справа*)

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.

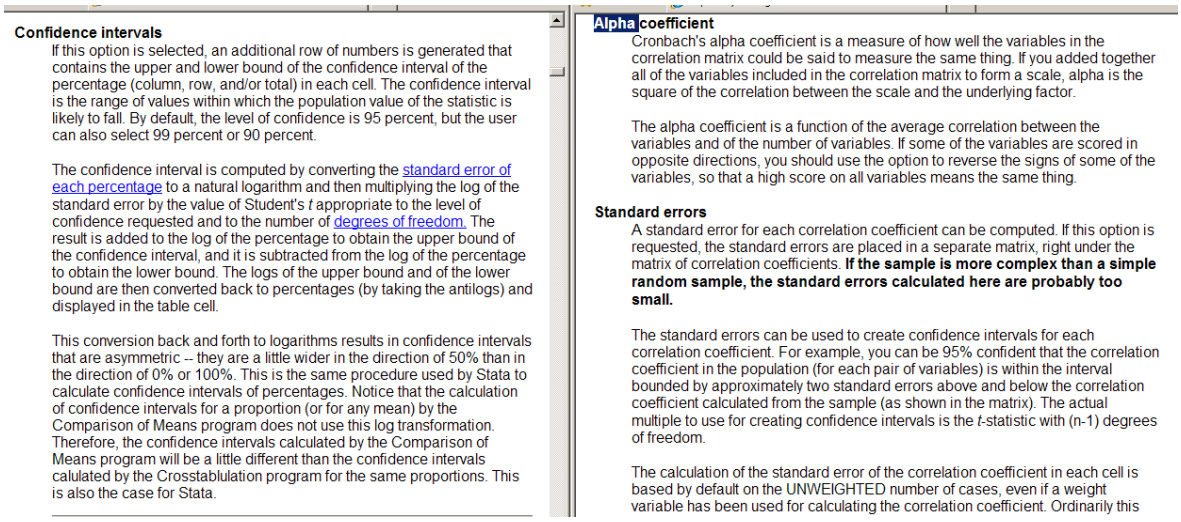


Рис. 7 Разделы справочной системы с описанием опции «Доверительный интервал» (Confidence Interval) (*слева*) и «Коэффициент Альфа» (*справа*)

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.

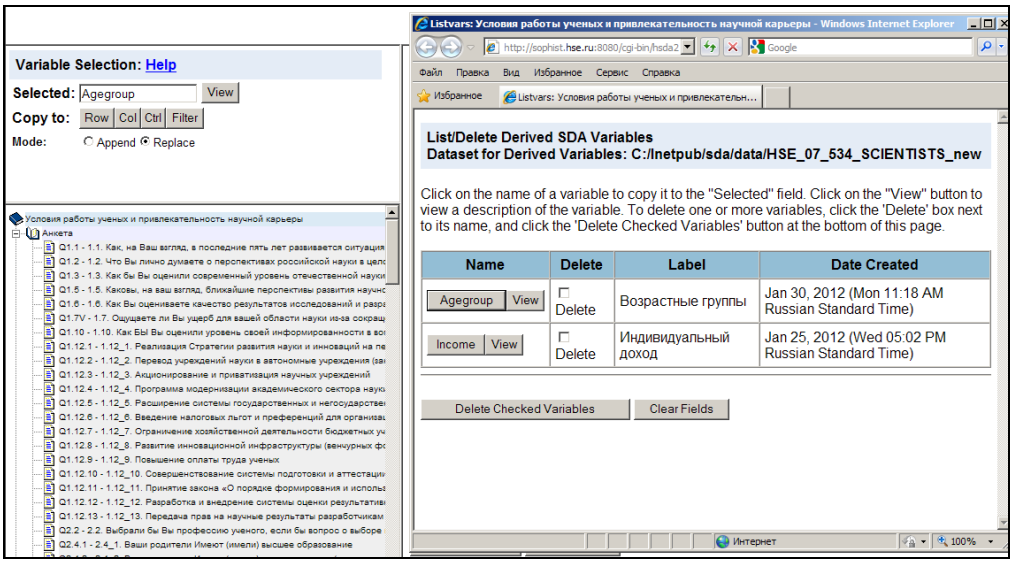


Рис. 8 Фреймы «Выбор переменных» и «Дерево переменных (*слева*) и окно «Распечатать/удалить построенные SDA переменные» (*справа*)

Подготовленные переменные отображаются не в дереве переменных, а в отдельном окне переменных, подготовленных пользователем – «Распечатать/удалить построенные SDA переменные» (List/Delete Derived SDA Variables. 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» (Analisys) 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.

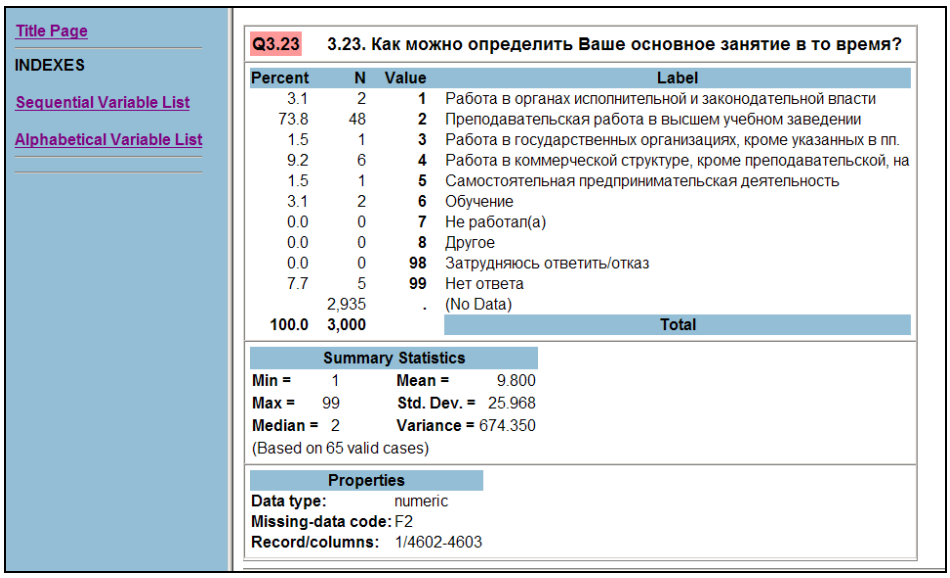
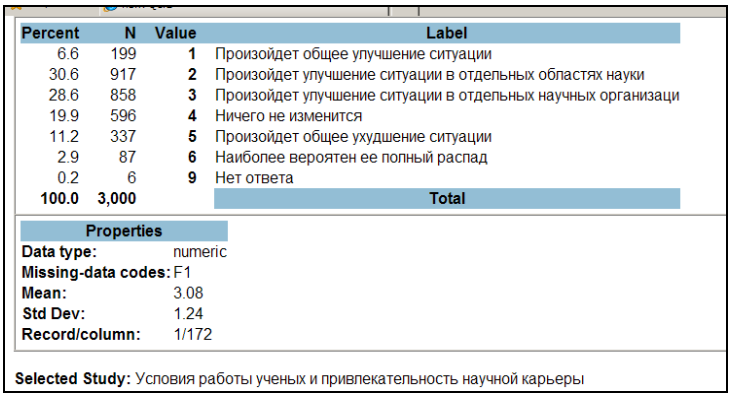


Рис. 9 Окно коудбука

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).

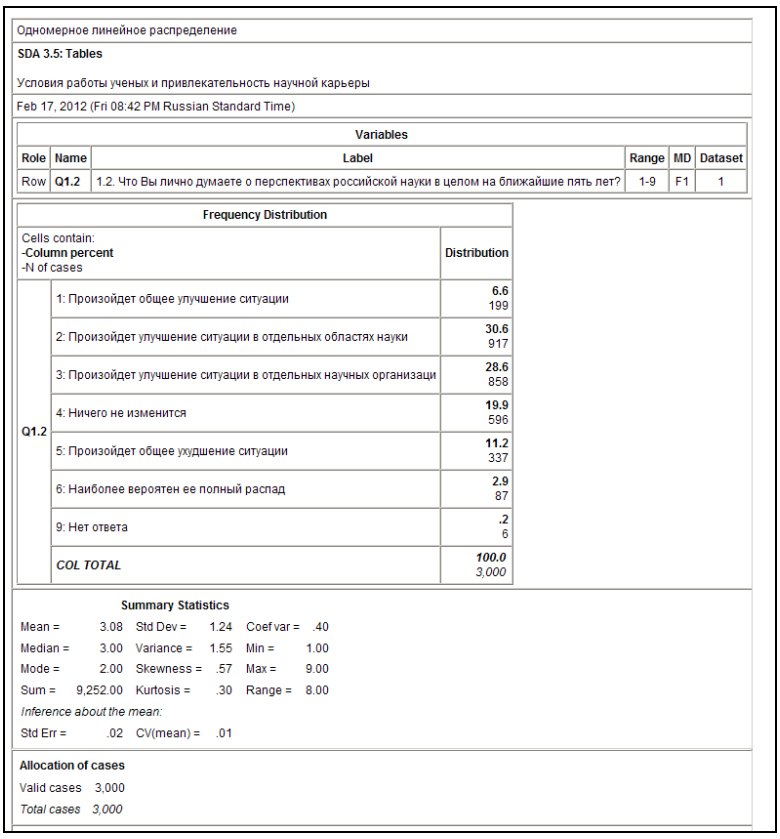


Percent	N	Value	Label
6.6	199	1	Произойдет общее улучшение ситуации
30.6	917	2	Произойдет улучшение ситуации в отдельных областях науки
28.6	858	3	Произойдет улучшение ситуации в отдельных научных организаци
19.9	596	4	Ничего не изменится
11.2	337	5	Произойдет общее ухудшение ситуации
2.9	87	6	Наиболее 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

Рис. 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: -**Column percent** -N of cases	**Distribution**
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

Inference about the mean:

Std Err =	.02	CV(mean) =	.01

Allocation of cases

Valid cases 3,000

Total cases 3,000

Рис. 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.

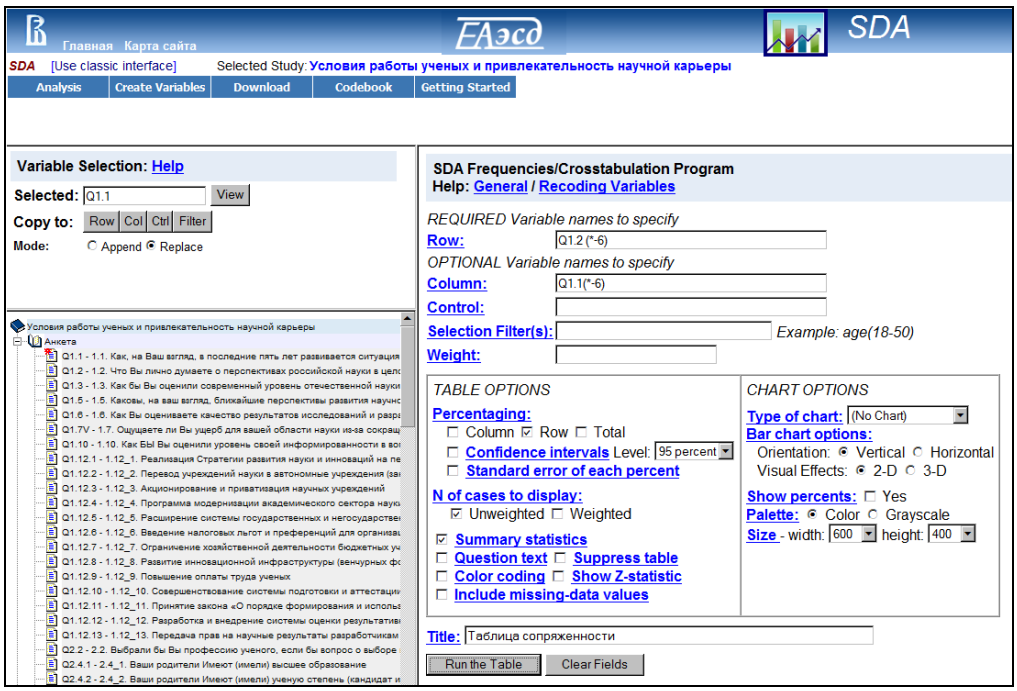
2.4.2 Построение таблиц сопряженности.

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 наблюденного числа случаев для каждой клетки таблицы (N of cases to display) как взвешенного, так и не взвешенного (для отображения взвешенных значений разумеется в соответствующем поле «Взвешивание» (Weight) должен быть указан весовой коэффициент.);
- рассчет процента по строке (Row),
- рассчет процента по столбцу (Column)
- рассчет табличного процента (Total),
- рассчет стандартной ошибки среднего (Standart errors of each percent),
- рассчет 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).



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

Cells contain:
-Row percent
-N of cases

		Q1.1						
		1 Наблюдается значительное улучшение ситуации	2 Наблюдается некоторое улучшение ситуации	3 Сохраняется стабильная ситуация	4 Сохраняется кризисная ситуация	5 Наблюдается некоторое ухудшение ситуации	6 Наблюдается значительное ухудшение ситуации	*ROW TOTAL*
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 организаци	**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*

Рис. 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).

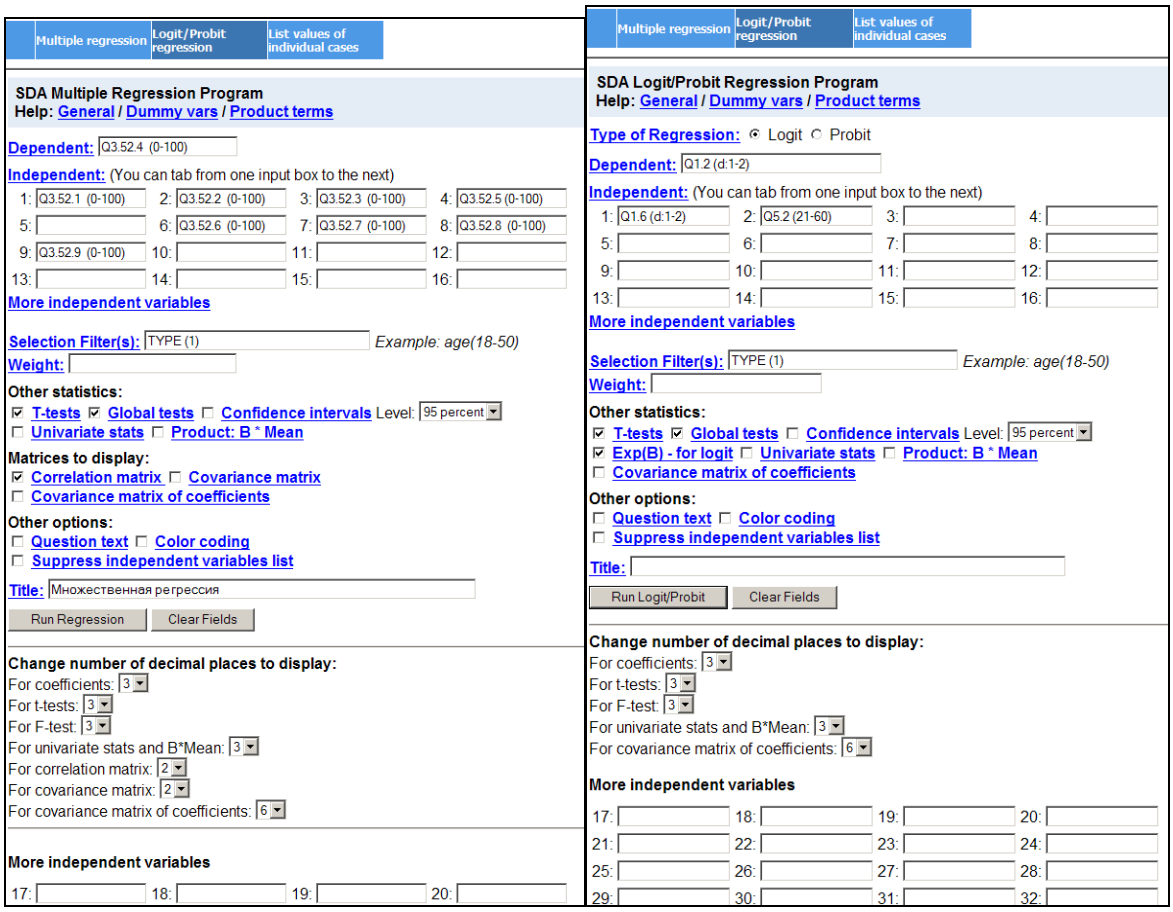


Рис. 13 Часть фрейма «Процедура» для процедур анализа «Множественная регрессия» (*слева*), и «Логит\Пробит регрессия» (*справа*)

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 вверху 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:

- «T-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**

Рис. 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*

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

3. ЗАКЛЮЧЕНИЕ

Рассмотренный программный комплекс обладает рядом следующих особенностей, на которые мы хотим еще раз обратить внимание:

1. Важнейшее преимущество данного программного комплекса (в сравнении, скажем, с пакетами статистического анализа, устанавливаемых на локальный компьютер пользователя) 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.