

РУКОВОДСТВО

ПОЛЬЗОВАТЕЛЯ

OCA New Line Technology



Киев

Бюро статистического анализа 2006

Содержание

Вступление	3
Структуры данных	4
Дневники.....	4
Трекинг	6
Специальные настройки паспорта	8
Подготовка и структура данных в OCA New Line	12
Описание проекта	13
Создание проекта для OCA New Line (программа DataMaker).....	15
Функция «Конвертор SPSS -> OCA»	16
Функция «Реструктурирование данных»	17
Создание дневниковых массивов	17
Проверка дневников	18
Функция «Создание дерева признаков»	18
Особенности работы с большими массивами данных.	18

Вступление

Данное руководство является документацией по подготовке результатов социологических и маркетинговых исследований в формате OCA New Line. OCA New Line является программой, позволяющей проводить статистический анализ данных, строить электронные отчеты. Все данные в этой среде объединяются в так называемые **проекты**.

Проект - это один или несколько массивов, имеющих общий паспорт. Также к проекту относятся построенные в нем описатели таблиц и целевые аудитории. Напомним, что таблицы в OCA New Line сохраняются в виде описания задания к построению таблицы, а не в виде результатов расчета, поэтому любая из таблиц может быть легко и быстро пересчитана для любой целевой аудитории.

Таблицы объединяются в так называемый электронный отчет, который представляет собою двухуровневую древовидную структуру описателей таблиц, сгруппированных в разделы.

Руководство рассчитано на людей, умеющих работать в среде программ OCA и OCA New Line.

Структуры данных

Традиционно данные социологических и маркетинговых исследований представляются в виде таблицы, где каждый столбец является признаком, а каждая строка – анкетой. Такое представление характерно для программ OCA, SPSS, SAS и т.п. OCA New Line не является исключением, так как работает с форматом пакета OCA и поддерживает аналогичный формат, но с некоторыми дополнительными возможностями.

Дневники

Первой такой возможностью является так называемый режим дневников.

Режим дневников позволяет организовать работу с анкетой имеющей, плавающую, не стабильную длину. Рассмотрим пример. Мы просим респондента ответить на вопрос, какие газеты он читает, а потом просим охарактеризовать каждую из них. При этом мы не знаем, сколько газет назовет респондент. Кто-то назовет одну, а кто-то пять. В анкете в таких случаях оставляют место для того, чтобы интервьюер записывал о каких именно газетах будет идти речь. Характеристика каждой газеты записывается как отдельный признак, количество которых заранее определяется с некоторым запасом. При анализе мы имеем сразу несколько признаков, характеризующих газеты, что неудобно. Нельзя построить одну общую таблицу - необходимо каким-то образом сводить наборы таблиц вместе. Кроме того, размер анкеты сильно вырастает из-за того, что необходимо иметь сразу несколько признаков для фиксации характеристик газет.

Подобные структуры могут быть вложенными одна в другую, что еще больше затрудняет работу с ними.

В качестве выхода из таких ситуаций в технологии OCA New Line предлагается разбивать такие анкеты на несколько анкет-записей (по количеству объектов – в случае нашего примера по количеству характеризующих газет). Эти несколько анкет являются одинаковыми во всем, кроме признаков, связанных с объектами исследования (в случае нашего примера с газетами).

Рассмотрим структуру данных наглядно. Так выглядит простое, «линейное» представление данных для нашего примера с газетами.

Номер анкеты	Пол	Газета 1	Х-ка газеты 1	Газета 2	Х-ка газеты 2
1	М	«Новости»	хорошо	«Известия»	отлично

2	Ж	«Правда»	хорошо	Нет Ответа	Нет Ответа
3	М	Нет Ответа	Нет Ответа	Нет Ответа	Нет Ответа

А так эти же данные будут выглядеть в дневниковом варианте.

Номер анкеты	Пол	Газета	Х-ка газеты
1	М	«Новости»	хорошо
1	М	«Известия»	отлично
2	Ж	«Правда»	хорошо
3	М	Нет Ответа	Нет Ответа

Как мы видим, анкета 1 как бы раздвоилась (ведь в ней оценивалось сразу две газеты). Такое представление является более удобным и компактным – есть один признак, содержащий газеты и один признак, содержащий их характеристики. Признак, характеризующий дневник, это в нашем случае признак «номер анкеты» - по нему можно понять, сколько анкет есть на самом деле, и какая запись к какой анкете относится. Этот признак, а точнее его номер, и необходимо указать в качестве параметра для дневникового режима при описании проекта (подробнее см. раздел «Описание проекта»). По нему программа будет определять и склеивать отдельные записи в один общий дневник.

Однако, как показывает практика, для фиксации интервьюером данных и ввода их оператором удобнее использовать обычный («линейный») формат. Он более понятен для оператора ввода и поэтому его использование избавляет от многих ошибок. Поэтому в пакет программ OCA New Line Technology входит специальная программа-конвертор **Svm.exe**. Эта программа позволяет автоматически преобразовать данные из «линейного» формата в дневниковый (подробнее об этой программе, с детальными примерами свертки, читайте в соответствующем разделе).

Теперь рассмотрим логику работы с дневниковыми проектами. Как мы видим, есть некий признак, по которому мы «свернули» наши данные в дневник (в нашем примере это название газеты), есть некоторые признаки, описывающие анкету-дневник в целом (в нашем примере пол). Понятно, что при построении частотных таблиц для нашего примера, количество мужчин должно быть равно 2, женщин 1, а различных читаемых газет 4 (включая вариант Нет Ответа)

При построении **частотных** таблиц действует правило: если две анкеты имеют одно и то же значение признака и одинаковый номер дневника, то вторая из них игнорируется.

При расчете сумм и среднего: суммируются (а для среднего - учитываются в базу) все анкеты (ничто не игнорируется). Каждая анкета-запись считается независимой анкетой.

ВНИМАНИЕ! При построении целевых аудиторий для проектов, которые имеют дневниковую структуру, отбираются только те записи внутри анкеты, которые удовлетворяют условию. Таким образом, как бы, часть анкеты будет отобрана, а часть нет.

Трекинг

В ОСА New Line реализована возможность работы с данными трекинговых (мониторинговых) проектов. Трекинговыми, считаются проекты, которые проводятся регулярно (ежегодно, ежеквартально, ежемесячно, еженедельно, ежедневно), по одним и тем же правилам, с использованием одинакового инструментария. Часто в них опрашиваются одни и те же самые люди (тогда такие проекты называются панельными). Исследование за один период времени (год, квартал, месяц, неделя, день) часто называют волной. При этом при анализе данных нескольких волн часто используют процедуру усреднения. Например, если человек в первой волне сказал что его зарплата 100 у.е., а во второй сказал, что ее повысили до 120 у.е. то зарплата за все трекинговое исследование (2 волны) будет $(100+120)/2=110$ у.е. Для получения подобного эффекта можно использовать взвешивание, однако это бывает не всегда удобно, так как веса необходимо рассчитывать с появлением каждой новой волны заново, а также для каждой целевой аудитории (ведь если отобрать в целевой аудитории только вторую волну то зарплата в нашем примере должна быть равна 120 у.е.). Также перерасчета весов требует и переход с одного уровня на другой (например, если исследование проводится еженедельно, его можно оценивать на уровне кварталов, месяцев и т.д.).

Для решения этой проблемы в технологии ОСА New Line используется режим трекинга. Для этого создается специальный признак, который определяет номер объекта (например, респондента) в трекинговом исследовании. Этот номер за ним закрепляется на протяжении всего исследования, а признак, в котором он хранится, описывается для данных ОСА New Line как трекинговый (подробнее см. раздел «Описание проекта»). Программа автоматически при подключении этого режима будет производить необходимое перевзвешивание. В отличие от режима дневников, режим трекинга, при некоторых настройках, может управляться пользователем программы ОСА New Line. Пользователь может включать и выключать его, выбирать на каком уровне трекинга он будет работать. (О настройках трекингов см. раздел «Описание проекта»).

В проектах OCA New Line волны одного трекингового исследования удобно представлять в виде отдельных массивов с одинаковым паспортом. Тогда человек, при загрузке проекта, сможет отбирать к работе только те волны, которые его интересуют.

Режим трекинга и дневника являются независимыми друг от друга и могут использоваться, при необходимости, как вместе, так и по отдельности. При совместном использовании следует помнить, что один дневник должен иметь одно и тоже значение признака, описывающего трекинг (не может половина дневника принадлежать одной волне трекинга, а вторая половина второй).

Специальные настройки паспорта

OCA New Line поддерживает некоторые специальные возможности для описания данных в паспорте. Данные описатели помещаются в комментарии к признаку.

@&EQUALITY X=Y, где X – номер альтернативы, Y – произвольное значение. При наличии такого описателя для расчетов статистик (таких как среднее, минимумы, максимумы и т.п.) будет браться не значение равное коду альтернативы, а то, равенство которому мы приписали. Это позволяет, например, работать с порядковыми шкалами, включающими отрицательные значения.

Пример.

Оцените по шкале от -3 до 3 Вашу удовлетворенность жизнью.

Оцените по шкале от -3 до 3 Вашу удовлетворенность жизнью.

О

1 -3

2 -2

3 -1

4 0

5 1

6 2

7 3

@&EQUALITY 1= -3

@&EQUALITY 2= -2

@&EQUALITY 3= -1

@&EQUALITY 4= 0

@&EQUALITY 5= 1

@&EQUALITY 6= 2

@&EQUALITY 7= 3

/

@&LINK_TO X>Y, где X – номер альтернативы, Y – номер признака с метрической шкалой. Этот описатель позволяет при расчете таблиц подменять вес ответа респондента. То есть, если для расчета частотной таблицы по признаку обычно используется формула $1 * \text{Вес}$, то в этом случае формула будет выглядеть как **значение_признака_Y * Вес**. При этом каждая альтернатива может ссылаться на свой признак, содержащий вес. Если Вам необходимо, чтобы все альтернативы признака ссылались на один и тот же признак, следует в комментарии записать следующую строку $* > Y$, где Y – номер признака с метрической шкалой.

Использовать теги @&LINK_TO и @&EQUALITY можно в порядковых шкалах, номинальных и шкалах с совместимыми альтернативами.

Пример 1.

Какой журнал Вы читает наиболее часто?

Какой журнал Вы читает наиболее часто?

j

1 Журнал 1

2 Журнал 2

3 Журнал 3

@&Link_To 1 > 10

@&Link_To 2 > 11

@&Link_To 3 > 13

/

Пример 2.

Какой журнал Вы читает наиболее часто?

Какой журнал Вы читает наиболее часто?

j

1 Журнал 1

2 Журнал 2

3 Журнал 3

@&Link_To * > 10

/

Технология OCA New Line поддерживает новый вид шкалы – каталог. Фактически, это метрическая шкала, использующаяся как номинальная. Так как при использовании каталогов в качестве кодов можно использовать любые числа (даже дробные и отрицательные), то использовать их удобно в том случае, если оператор вводит информацию о неких категориях, которых либо очень много, либо которые кодируются по некоторым специальным правилам. Например, код товара равен код производителя*1000+код ассортиментной единицы.

При построении таблиц и т.п. для списков работают те же правила что и для номинальных шкал.

Для описания списка необходимо в комментарии к метрической шкале указать ключ @CATALOGUE и **после** него описать элементы списка в формате

X1 # N1

X2 # N2

...

где X1 значения (коды), а N1 наименования (описания кодов).

Пример.

Наименование товара.

Наименование товара.

m

@CATALOGUE

1 # Товар 1

15 # Товар 2

100 # Товар 3

18 # Товар 4

/

Если есть несколько признаков-каталогов, имеющих одинаковый кодировочный список, список можно хранить только в одном из них. Из остальных можно на него сослаться используя директиву

@TAKE_CATALOGUE_FROM=X

где X номер или имя признака, из которого следует взять список.

Подготовка и структура данных в OCA New Line

Все данные находятся в стандартной папке DATA, которая находится в каталоге, в котором установлена OCA New Line.

Поддерживается до 500 проектов одновременно и до 500 массивов к каждому проекту.

Каждый проект находится в каталоге (он может называться произвольно).

В каталоге находятся:

1. Файл “project.txt”. В нем – первая строка название проекта, далее идет описание проекта.
2. Паспорт проекта (*.dcl)
3. Древоподобное описание паспорта (имя_паспорта.tre и имя_паспорта.trd)
4. Массивы (имя_массива.frm) – один или несколько.
5. Описания к массивам (имя_массива.txt). Формат описания массива аналогичен описанию проекта (первая строка название массива – далее описание).
6. Файл project.ini – настройки проекта. (см. Далее)
7. Каталог “Tables”, где хранятся описания к таблицам. Файлы-описания имеют расширение *.TBL и содержат описания в стандарте Ttable. Первая строка файла – категория, к которой относится таблица, вторая строка – название таблицы. Каталог может содержать до 10000 файлов-описаний.
8. Каталог «Filtr» (для фильтров). Он может быть пустым, но должен быть обязательно.

ВНИМАНИЕ! В одном проекте может быть несколько массивов, но все они должны иметь **один общий** паспорт. Файл с расширением *.DCL в каталоге каждого проекта должен быть **один**.

ВНИМАНИЕ! В массиве не должно быть двух признаков с одинаковыми именами. В одном признаке не должно быть двух альтернатив с одинаковыми описаниями к ним. При построении дерева Вас попросят переименовать такие совпадения.

ВНИМАНИЕ! Нельзя в названиях признаков и альтернатив использовать двойные кавычки и символы <>. При создании проекта с помощью программы DataMaker вышеупомянутые символы будут автоматически заменены: “ на ‘, <> на [] .

Описание проекта

Описание проекта содержится в файле **project.ini**. Этот файл необходимо заполнить аккуратно и внимательно, так как при некорректном его заполнении могут возникнуть серьезные ошибки.

Он может содержать следующие параметры:

Universal=X, где X объем генеральной совокупности, для которой проводилось исследование. Если этот параметр не указать, объемом генеральной совокупности будет считаться число, равное количеству анкет в массивах проекта.

Weigh=X#Y, где X признак (обязательно метрический) для взвешивания. Если этот параметр не указывать, взвешивание будет заблокировано. Y – текст описание взвешивания. Количество разных весов – не более 3. Каждый описывается своей отдельной строкой.

SetWeigh - если указать этот флажок, то при загрузке проекта будет включаться вес (он должен быть описан, если несколько – будет работать первый по вхождению).

SuperWeigh – если указать этот флажок, то при загрузке проекта будет включаться вес (он должен быть описан, если несколько – будет работать первый по вхождению) и отключить его не будет возможности.

DairyParams=X1, где X1 признак (обязательно метрический), определяющий режим дневников. Указывается только в случае когда данные подготовлены для работы этом режиме.

PanelParams=X1#Y, где X1 признак (обязательно метрический), определяющий режим трекинга, Y - текст описание панели. Всего может быть до 10 штук разных трекинговых настроек. Каждая описывается своей отдельной строкой. Автоматически подключается первая по вхождению. Указывается только в случае когда данные подготовлены для работы этом режиме.

NoPanelOnStart – флажок, отключающий при загрузке проекта трекинговые настройки (они могут быть подключены при желании пользователем в программе OCA NL).

SuperPanel - если указать этот флажок, то у пользователя не будет возможности менять настройки трекинга в проекте.

NoPanel=X1 [X2 X3 ...], где X1 X2 X3 - номера признаков, присутствие которых в любой расчетной таблице, автоматически отключает режим трекинга. Это необходимо для признаков, по которым рассчитывается трекинг. Например, у нас есть данные за 2 месяца

(заказан месячный трекинг): 800 объектов (анкет) по 400 собранных каждый месяц. При построении обычных таблиц мы будем иметь дело с 400 анкетами. Однако при просмотре распределения по месяцам (признак X1), объектов будет 800.

DSorted - для массивов состоящих из дневников. При настройке на такой проект автоматически происходит сортировка массива по признаку, определяющему номер дневника. Если массив большой, этот процесс может занять много времени. В таком случае можно заранее отсортировать массив и прописать ключ DSorted. Тогда при загрузке сортировка не проводится.

<List>

N1

N2

...

</>

Этот параметр позволяет задать порядок показа массивов, относящихся к проекту в списке. N1, N2 - имена массивов (без пути). Каждый массив записывается с новой строки.

ПАРАМЕТРЫ ДЛЯ МЕДИАПЛАНИРОВЩИКА (только для медиапроектов)

MP_Source=X1 [X2 X3]. Список номеров признаков, в которых описаны носители рекламы. Признак НЕ должен быть метрическим.

MP_Time=X1 [X2 X3]. Список номеров признаков, в которых описаны временные параметры рекламы. Признак НЕ должен быть метрическим.

MP_Period=X1 [X2 X3]. Список номеров признаков, в которых описаны периоды рекламы (будни, выходные и т.п.). Признак НЕ должен быть метрическим и не должен быть с совместимыми альтернативами.

Создание проекта для OCA New Line (программа DataMaker)

Для того, чтобы оформить данные для последующего их использования в среде OCA New Line, воспользуйтесь программой DataMaker. Запустив ее, и следуя инструкциям, Вы легко сможете создать и описать свой проект, сформировав в конце архив своего проекта в виде файла с расширением *.onle. Позже, воспользовавшись функцией «Импорт данных», Вы сможете импортировать из этого файла данные в OCA New Line.

При этом Вы можете защитить свои данные паролем. Тогда импортировать их сможет только тот, кто знает пароль.

С помощью программы DataMaker Вы можете не только создавать новые проекты для OCA New Line, но и редактировать настройки готовых проектов.

Функция «Конвертор SPSS -> OCA»

Данная функция предназначена для конвертации массивов SPSS (*.sav) в массив OCA (*.frm).

Для экспорта необходимо:

1. Нажав на кнопку "Загрузить массив", указать экспортируемый массив SPSS (*.sav).
2. В текстовом окне появится информация о загруженном массиве.
3. Нажав на кнопку "Экспорт в OCA" указать имя будущего паспорта массива (*.dcl). Массив данными будет называться также как и паспорт, но с расширением *.frm

При экспорте действуют следующие правила:

1. Текстовые признаки (не Numeric) при экспорте игнорируются.
2. Если нет описанных меток к значениям или все значения с метками соответствуют значениям, описанным как missing - шкала признака определяется как метрическая.
3. Если есть описанные метки к значениям, значения целые и находятся в диапазоне 0-254 - шкала признака определяется как порядковая.
4. В остальных случаях шкала признака определяется как каталог (тип шкалы поддерживаемый OCA New Line, в OCA for Windows интерпретируется как метрическая шкала).
5. Все missing value (system missing и значения, описанные как missing value) перекодируются в НеОтвет.
6. В качестве названия признака используется значение поля Name. В качестве текста вопроса используется поле Label (если оно пустое используется значение поля Name). Значение поля Name используется также в комментариях как имя признака (значение директивы @qname).

Функция «Реструктурирование данных»

Эта функция предназначена для свертки (перегруппировки и перекодировки) данных разного типа.

Создание дневниковых массивов

На главной форме находятся настройки для свертки дневниковых данных (подробнее см. раздел «Структуры данных»). Для работы необходимо, нажав на кнопку «Обзор», указать массив, который будет сворачиваться. Потом в поле «Начало блока 1» необходимо указать номер признака, являющегося первым в блоке свертки. В поле «Конец блока 1», соответственно, указывается номер последнего признака в блоке свертки. В поле «Количество блоков» указывается, сколько блоков необходимо свернуть. После заполнения этих параметров необходимо нажать клавишу «ОК» и свертка будет проведена. Новый массив будет создан в той же папке, где находится и старый. Его имя будет new1.frm с паспортом new1.dcl. Если такой массив уже есть в папке, будет создан массив new2.frm с паспортом new2.dcl и так далее.

Если дневниковые структуры являются вложенными одна в другую необходимо проводить свертку несколько раз, свернув сперва верхнюю структуру, потом ту, что ниже и так далее. Рассмотрим пример сложной многоуровневой свертки массива.

Массив.

#1#	#2#	#3#	#4#	#5#	#6#	#7#	#8#	#9#	#10#	#11#	#12#	#13#	#14#	#15#	#16#	#17#
1	М	А	б	3	в	5	а	5	В	а	3	\$	\$	\$	\$	30
2	Ж	А	в	4	а	3	\$	\$	\$	\$	\$	\$	\$	\$	\$	28
3	М	Б	б	5	в	7	а	4	Г	а	3	ж	3	б	5	42
4	М	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	45

Где:

#1# - номер анкеты

#2# - пол

#3# - название регулярно читаемой респондентом газеты 1

#4# - название любимого раздела в этой газете 1

#5# - оценка этого раздела в последнем номере

- #6# - название любимого раздела в этой газете 2
- #7# - оценка этого раздела в последнем номере
- #8# - название любимого раздела в этой газете 3
- #9# - оценка этого раздела в последнем номере
- #10# - название регулярно читаемой респондентом газеты 2
- #11# - название любимого раздела в этой газете 1
- #12# - оценка этого раздела в последнем номере
- #13# - название любимого раздела в этой газете 2
- #14# - оценка этого раздела в последнем номере
- #15# - название любимого раздела в этой газете 3
- #16# - оценка этого раздела в последнем номере
- #17# - возраст

В массиве отказ отвечать мы обозначим символом \$.

Итак, как мы видим, здесь необходима двухуровневая свертка (на уровне газеты и на уровне раздела). При этом верхним является уровень газеты. С него мы и начнем. Указываем как «Начало блока 1» - 3. Как «Конец блока 1» указываем 9. Количество блоков у нас 2. После этой процедуры мы получим массив new1.frm с паспортом new1.dcl, который будет содержать следующее:

#1#	#2#	#3#	#4#	#5#	#6#	#7#	#8#	#9#	#10#
1	М	А	б	3	в	5	а	5	30
1	М	В	а	3	\$	\$	\$	\$	30
2	Ж	А	в	4	а	3	\$	\$	28
3	М	Б	б	5	в	7	а	4	42
3	М	Г	а	3	ж	3	б	5	42
4	М	\$	\$	\$	\$	\$	\$	\$	45

Где:

#1# - номер анкеты

#2# - пол

#3# - название регулярно читаемой респондентом газеты

#4# - название любимого раздела в этой газете 1

#5# - оценка этого раздела в последнем номере

#6# - название любимого раздела в этой газете 2

#7# - оценка этого раздела в последнем номере

#8# - название любимого раздела в этой газете 3

#9# - оценка этого раздела в последнем номере

#10# - возраст

Анкеты 1 и 3 «раздвоились», так как содержат информацию о двух газетах.

Теперь можно провести свертку по следующему уровню. Указываем как «Начало блока 1» -

4. Как «Конец блока 1» снова указываем - 5. Количество блоков у нас равняется 3.

Получим массив new2.frm с паспортом new2.dcl, который будет содержать следующее:

#1#	#2#	#3#	#4#	#5#	#6#
1	М	А	б	3	30
1	М	А	в	5	30
1	М	А	а	5	30
1	М	В	а	3	30
2	Ж	А	в	4	28
2	Ж	А	а	3	28
3	М	Б	б	5	42
3	М	Б	в	7	42
3	М	Б	а	4	42
3	М	Г	а	3	42
3	М	Г	ж	3	42
3	М	Г	б	5	42
4	М	\$	\$	\$	45

Где:

#1# - номер анкеты

#2# - пол

#3# - название регулярно читаемой респондентом газеты

#4# - название любимого раздела в этой газете

#5# - оценка этого раздела в последнем номере

#6# - возраст

Это и будет наш конечный свернутый массив. В качестве дневникового параметра у него будет выступать 1 (первый признак – «Номер анкеты»).

Проверка дневников

После создания дневникового массива имеет смысл проверить его. Для этого служит функция «Файл» -> «Проверка дневников». В данном случае, под проверкой подразумевается анализ некоторых признаков на уникальность их значения внутри одного дневника. Как мы помним, повторяющиеся значения внутри дневника игнорируются при некоторых расчетах (подробнее см. «Структуры данных»). Конечно, многие признаки могут и будут повторяться внутри дневника (в нашем предыдущем примере это, например, пол респондента – для всего дневника он будет конечно же одинаковый). Однако признак, по которому мы проводили свертку, повторяться не должен. Это не является ошибкой, однако часто указывает на ошибки ввода. В нашем предыдущем примере таким признаком можно считать сочетание признаков «название регулярно читаемой респондентом газеты» и «название любимого раздела в этой газете» (поскольку свертку мы проводили в 2 этапа). Ведь не мог человек более одного раза оценивать один и тот же раздел одной и той же газеты.

Для запуска проверки необходимо указать массив, признак, определяющий дневники («Номер дневника») и признак по которому проводить проверку. После этого следует нажать на кнопку «Старт». Сообщения о найденных ошибках и неточностях будут выведены в текстовое поле внизу формы. Настройка «Помечать проблемные анкеты НеОтветом в признаке» дает возможность прописывать указанный Вами признак НеОтветами в анкетах, которые содержат ошибки (потом их легко можно отобрать фильтром). То есть, если есть повтор значения признака, который мы проверяем внутри одного дневника, то для анкеты, которая содержит второе значение, в указанном признаке будет поставлен НеОтвет.

Если указать номер признака 0, ничего прописываться не будет.

Создание номинальных признаков с совместимыми альтернативами с длинным списком альтернатив (более 250)

В некоторых случаях бывает необходимость представить признак в виде шкалы с совместимыми альтернативами количеством более 250. Объединение нескольких признаков, с разными кодировочными списками, в один описывается в разделе «функция «Создание дерева признаков».

Однако корректно создать такой набор признаков достаточно сложно технически. Функция «Файл» -> «Свернуть в j-шкалу» позволяет делать это автоматически из нескольких признаков-каталогов. Во время ввода анкет создаются несколько признаков-каталогов, в которые поочередно вводятся ответы респондента. После этого, указав массив, список

признаков-каталогов, в которые вводились ответы, и название нового признака, который объединит в себе все ответы, нажмите кнопку «ОК». Признаки будут корректно перекодированы в необходимое количество признаков с совместимыми альтернативами, (они будут дописаны в конец массивами). Первый из них будет иметь указанное Вами имя, остальные служебное имя NO NAME. После этого при построении дерева признаков с помощью функции «Создание дерева признаков» можно объединить кодировочные списки этих признаков в один. При этом проследите, чтоб при формировании дерева была включена функция «Автоисключение» (чтобы были исключены служебные альтернативы, начинающиеся с двух подряд записанных точек - «..»).

Если отметить параметр «Учитывать варианты ответа с нулевой частотой», то в кодировочный список создаваемых признаков попадут как все описанные в каталоге варианты ответов, так и все введенные в качестве ответов варианты (для них в качестве интерпретации будет записано их значение). Если вышеуказанный параметр не отмечен, то в кодировочный список не попадут, описанные в каталоге, но неиспользуемые в анкетах варианты ответов.

Функция «Создание дерева признаков»

Эта функция необходима для группирования признаков паспорта в формате OCA (*.dcl). Она группировать признаки в древообразные структуры для последующей работы с ним в OCA New Line.

Слева на форме находится контейнер, в котором пользователь формирует желаемую древообразную структуру паспорта.

Кнопка «Новый корень» - создает новую группу в корне дерева, с которым ведется работа.

Кнопка «Новая группа» - создает новую группу, которая является подгруппой активной (выделенной) группы.

Для переименования любого элемента дерева необходимо щелкнуть по нему мышкой зажав при этом клавишу 'Ctrl'. Следует помнить, что переименования названий признаков или альтернатив изменяет не только текст отображение в дереве, но и соответствующие тексты в паспорте массива.

Нажав на кнопку с изображением стрелки, указывающей влево, можно «перенести» один или несколько выделенных признаков в какую-либо группу из правого списка. Признак представляет собой группу, элементы которой – альтернативы. Для выделения сразу нескольких признаков, действуют стандартные правила Windows. Можно, зажав кнопку Ctrl, кликнуть мышкой на каждом из желаемых признаков. Либо, если они идут подряд, отметить первый и, зажав клавишу Shift, кликнуть на последнем.

Нельзя переносить признаки внутрь групп, которые представляют собой другие признаки.

Каждый признак может быть внесен в дерево один раз (дерево не может содержать более одной ветки относящейся к одному и тому же признаку). Признаки, которые не текущий момент еще не перенесены в дерево, или ранее были внесены, но после удалены из дерева, помечаются символом звездочки. С помощью двух кнопок с изображением синих стрелок вверх и вниз можно автоматически перейти к первому сверху (снизу, соответственно) не перенесенному в дерево признаку.

В качестве названия ветки представляющей признак в дереве можно использовать либо название, либо текст вопроса. Для настройки этого параметра служит набор переключателей «Для названия ветки дерева использовать». Также есть третий вариант: «то что длиннее». В

случае его выбора для подписи будет использован тот текст, который длиннее, а значит, часто более информативен. Если длинна названия признака и текста вопроса совпадает, берется текст вопроса. Этот вариант задан по умолчанию.

Кнопка «Удалить» - удаляет выделенный элемент дерева. Удалять можно ветки представляющие собой признаки или их группы.

Кнопки с изображением стрелок «вверх» и «вниз» позволяют перемещать группы, признаки и альтернативы, менять их между собой местами.

Если признак имеет название NO NAME – он является частью кодировочного списка другого признака. Это значит, что здесь и в дальнейшем его альтернативы являются продолжением списка альтернатив того признака, к которому он «подшивается». Сшивать можно признаки, имеющие шкалу с совместимыми альтернативами.

ВНИМАНИЕ! Помните, что сшиваемые признаки должны иметь одинаковую структуру ответов и НеОтветов. То есть если хоть в одном из них есть ответ, то ответ должен быть и во всех остальных. Для этого, в каждом из соединяемых признаков создается альтернатива «Есть ответ». Эта альтернатива должна быть введена в каждом из сшиваемых признаков, если хотя бы в одном из них есть ответ. При оформлении дерева эта альтернатива выносится в список исключаемых альтернатив, являясь, таким образом, невидимой для конечного пользователя.

При группировании необходимо отметить тот признак, частью которого является группируемый признак, и после этого нажать кнопку с изображением стрелки, указывающей влево. Иным способом внести признак с названием NO NAME в дерево невозможно.

Иногда бывает необходимо скрыть некоторые неиспользуемые альтернативы. Для этого перед переносом признака в дерево необходимо указать список кодов таких альтернатив в строке «Исключить». Если необходимо исключить несколько альтернатив, список их кодов следует задавать через пробел, для идущих подряд кодов можно использовать тире (запись «1-3» эквивалентно записи «1 2 3»). После переноса признака список «Исключить» автоматически очищается. Если отметить флажок «Автоисключение», то поле «Исключить» будет автоматически заполняться при переносе признака номерами альтернатив, чье описание начинается с «..» (без кавычек).

ВНИМАНИЕ! Во избежание недоразумений в дальнейшем следует помнить, что скрывать следует только «пустые», неиспользуемые альтернативы, так как скрытые в процессе создания дерева альтернативы, не будут выводиться не только в древообразном списке

признаков, но и будут отсутствовать в таблицах, будут игнорироваться при расчетах статистик и т.п.

При щелчке на любом из элементов дерева, внизу отображается его номер (№). Этот параметр необходим при подготовке данных для медиапланировщика (специальное приложение для медиаисследования – проектов измерения состава и размера аудитории СМИ - OCA New Line Media).

В поле «Инфо по проекту» необходимо указать некоторую краткую, описательную информацию о проекте (например, кто и для кого его делал). Позже, эта информация будет видна в подписях к таблицам, которые строятся в OCA New Line.

При нажатии на кнопку «Генерация», происходит генерация файлов сгруппированного паспорта. Перед генерацией проверяется:

1. Отсутствие признаков с одинаковыми именами.
2. Отсутствие альтернатив внутри одной группы-признака с одинаковыми именами.

Без выполнения вышеуказанных условий генерация невозможна. В случае нахождения ошибок пользователю выводится соответствующее сообщение. Если отметить флажок «Автоматическое исправление совпадений при генерации», то все совпадения будут исправлены автоматически (совпадающим объектам будут приписаны номера).

Пример.

До автоматического исправления:

Признак

Признак

После:

Признак

Признак (1)

В каталоге, где находится массив, во время генерации создаются два файла:

имя_паспорт.tre

имя_паспорт.trd.

Для проведения генерации совсем не обязательно чтобы дерево включало в себя все существующие признаки. Некоторые признаки можно скрыть от конечного пользователя (например, служебные признаки, веса и т.п.).

Следует помнить, что при изменении паспорта следует менять соответствующие измененным признакам элементы дерева.

Особенности работы с большими массивами данных.

1. При работе с дневниками следует помнить, что один дневник не должен состоять из более чем 30 000 анкет. При работе с большими массивами, состоящими из дневников, следует их отсортировать по признаку, определяющему номер дневника и поставить в project.ini ключ DSorted. Тогда эти массивы будут быстрее загружаться.

2. При работе с большими массивами данных можно создать в массиве признак следующего вида.

N

1

/

Значения этого признака во всех анкетах должны быть равны 1. Это позволит быстрее загружать массив при настройке на рабочий проект.

При создании дерева признаков программой TreeMaker, этот признак лучше не включать в список, чтобы он остался невидим конечному пользователю.