ТЕМА 11

АНАЛИЗ С ПОМОЩЬЮ CHAID

ОГЛАВЛЕНИЕ

1. CHAID АНАЛИЗ: ПОНЯТИЕ И НАЗНАЧЕНИЕ

1. CHAID анализ: понятие и назначение

CHAID (Chi Squared Automatic Interaction Detection) анализ – используется для построения прогностической модели, основанной на системе классификации. Это метод эффективного поиска взаимосвязи между предикторными переменными и категориальным откликом. Удобен в использовании, если имеется много потенциальных предикторов, которые сложно анализировать посредством таблиц сопряженности, из-за их большого количества.

Анализ подразделяет выборку на ряд подгрупп, которые:

- имеют сходные характеристики по отношению к конкретной переменной отклика
- максимизируют наши способности прогнозировать значения переменной отклика

Базовое отличие CHAID анализа от регрессионного заключается в том, что взаимосвязь между значением зависимой переменной и значениями независимых переменных представлена не в виде общего прогнозного уравнения, а в виде **древовидной структуры**, которую получают с помощью иерархической сегментации данных.

1. CHAID анализ: понятие и назначение

- **CHAID** – один из методов анализа с помощью дерева решений.
  (Остальные методы: **ECHAID, CRT, QUEST**)

- **CHAID** позволяет осуществлять многомерные расщепления узлов (в отличие от **CRT** и **QUEST**, где используется бинарное). Каждый узел при разбиении может иметь более 2 потомков, поэтому **CHAID** имеет тенденцию выращивать достаточно раскидистые деревья.

- По сравнению с другими методами, **CHAID** характеризуется умеренным временем вычислений.

- Помимо прочего, метод **CHAID** обладает собственным способом обработки пропущенных значений. Пропуски рассматриваются как отдельная фактическая категория.

1. CHAID анализ: понятие и назначение

- **Exhaustive CHAID (Исчерпывающий CHAID)** – является модификацией CHAID и позволяет обойти некоторые недостатки CHAID:
  - В алгоритме **CHAID** слияние категорий останавливается как только обнаруживается, что все оставшиеся категории статистически различны. **ExCHAID** продолжает сливать категории переменной-предиктора, пока их не останется только
  - **ExCHAID** сливает категории в разных комбинациях, для того, чтобы найти оптимальное разбиение на 2 категории для данной переменной-предиктора (такое, что p-значение будет минимальным)

- **CRT (Classification and Regression Trees)** – программа деревьев классификации, которая при построении дерева осуществляет полный перебор всех возможных вариантов одномерного ветвления (бинарный).

- **QUEST (Quick, Unbiased, Efficient Statistical Tree)** – это программа деревьев классификации, в которой используются улучшенные варианты метода рекурсивного квадратичного дискриминантного анализа и которая содержит ряд новых средств для повышения надежности и эффективности деревьев классификации, которые она строит. (бинарный).

1. CHAID анализ: понятие и назначение

- **Корневой узел** – верхний разбиваемый узел, представляющий всю выборку.
- **Узел-сын** – новые узлы, получившиеся в результате разбиения.
- **Узел-отец** – узел, который был расщеплен.
- **Терминальные узлы** – окончательные узлы, которые в дальнейшем не разбиваются. Их еще называют листьями, потому что в них рост дерева решений останавливается. Лист представляет собой наилучшее окончательное решение.
- **Глубина дерева** – количество уровней, образующихся от узлов, не считая родительский узел (можно установить вручную в SPSS).
- **Правило остановки** – критерий определения "подходящего размера" дерева классификации; состоит из глубины дерева, минимальное количество наблюдений в узле-отце и в узле-сыне.

В результате получаем **дерево решений** с ветвями – переменными предикторами, которые выделяют выборку в различные группы.

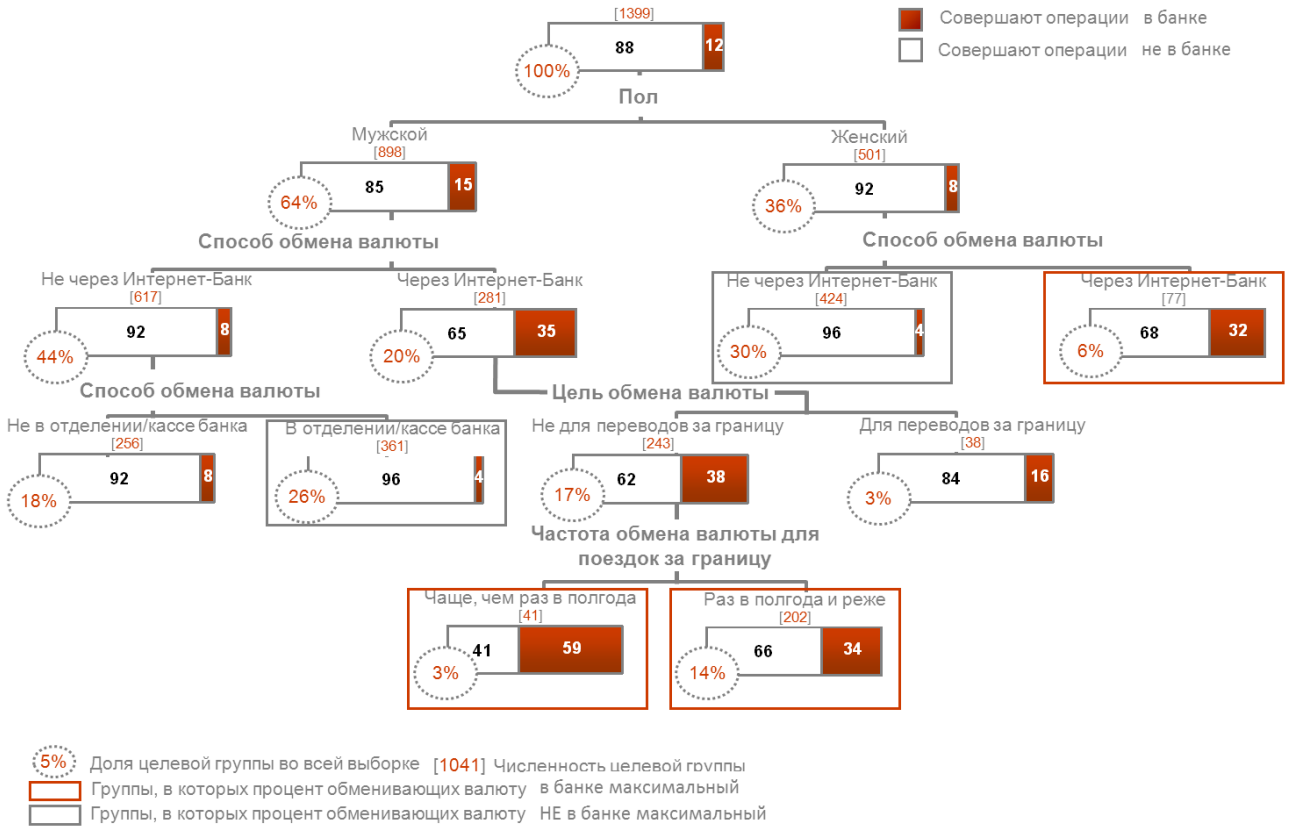
1. CHAID анализ: понятие и назначение

Дерево решений (decision tree) – это способ представления данных в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел, дающий решение. Под правилом понимается логическая конструкция, представленная в виде «если … то …».

Основная **задача** статистического анализа данных с помощью деревьев решений заключается в том, чтобы используя заданный набор наблюдений (называемый обучающей выборкой), уловить скрытые статистические закономерности в данных (как одни случайные (независимые) характеристики влияют на интересующую нас (зависимую) характеристику) и построить модель зависимости в виде дерева решений.

1. CHAID анализ: понятие и назначение

Деревья решений идеально приспособлены для **графического представления**, и поэтому сделанные на их основе выводы гораздо легче интерпретировать, чем если бы они были представлены только в числовой форме.



1. CHAID анализ: понятие и назначение

Преимущества:

- Метод работает **с переменными всех типов**, даже с номинальными (в отличие от других методов сегментации, в первую очередь, кластерного анализа)
- Широкая сфера применимости деревьев классификации делает их весьма привлекательным инструментом анализа данных
- Как метод разведочного анализа или как последнее средство, когда отказывают все традиционные методы, деревья классификации, по мнению многих исследователей, не знают себе равных

2. АЛГОРИТМ CHAID

Процедура анализа с помощью CHAID включает:

1. Поиск самого сильного фактора, который наибольшим образом объясняет различия.

2. Перебор всех заданных предикторов, поиск комбинаций значений и нахождение лучшего результата (который максимизирует различия). Выделение групп по найденному лучшему результату.

3. Повторение процесса (пунктов 1 и 2) с целью нахождения оптимального решения для второго уровня и т.д. для всех возможных уровней.

4. Представление результатов в виде дерева решений.

Exhaustive CHAID

Представляет собой модификацию **CHAID**, разработанную с целью устранения некоторых недостатков метода **CHAID**.

- При объединении категорий **Exhaustive CHAID** продолжает объединение наименее значимо различающихся категорий до тех пор, пока не останутся **только две категории** (**CHAID** останавливается, когда остаются только значимо различающиеся категории).
- Exhaustive **CHAID** выбирает разбиение, показывающее **наибольшую статистическую значимость**, из большего набора возможных разбиений.
- Exhaustive **CHAID** требует больше процессорного времени, но повышает шансы выбрать разбиения категорий, дающих наилучшие предсказания.

Поправки Бонферрони

При выполнении статистических тестов **CHAID** автоматически корректирует их уровни значимости для различных комбинаций категорий предиктора. Эти корректировки называются **поправками Бонферрони,** которые основываются на числе тестов и связаны с уровнем измерений предиктора.

Наличие поправок Бонферрони позволяет управлять уровнем ошибки первого рода.

Типы переменных

Каждую переменную можно охарактеризовать типом значений, которые она имеет, и тем, что эти значения измеряют. Эту общую характеристику называют **уровнем (типом) измерений переменной.** При анализе методом **CHAID** можно использовать тот тип переменной, который был задан в исходном файле данных SPSS, однако его можно изменить для нужд текущего анализа.

Тип предикторной переменной влияет на то, как при анализе методом **CHAID** будут объединяться категории, которые не покажут значимого различия.

2. Алгоритм CHAID

Можно задать три типа переменных:

- **Номинальный тип** – характерен для категориальных переменных с дискретными значениями, когда значениям не приписывается конкретный порядок. Можно объединять любые категории, если они не различаются значимо (например, номинальная переменная регион – любые регионы могут объединиться, если они не различаются значимо по целевой переменной).

- **Порядковый тип** – характерен для переменных с дискретными значениями, когда задан порядок значений. Две категории могут быть объединены, только если есть возможность присоединения к ним промежуточных категорий (например, люди с доходом меньше 30 т.р. можно объединить с теми, у кого доход выше 40 т.р., если к ним также отнести людей с доходом от 30 до 40 т.р.)

- **Непрерывный тип** – по умолчанию CHAID преобразует непрерывную числовую предикторную переменную (например, возраст в годах) в порядковую, имеющую 10 категорий с приблизительно равным числом наблюдений. Эти категории формируются путем объединения соседних значений с исходной переменной.

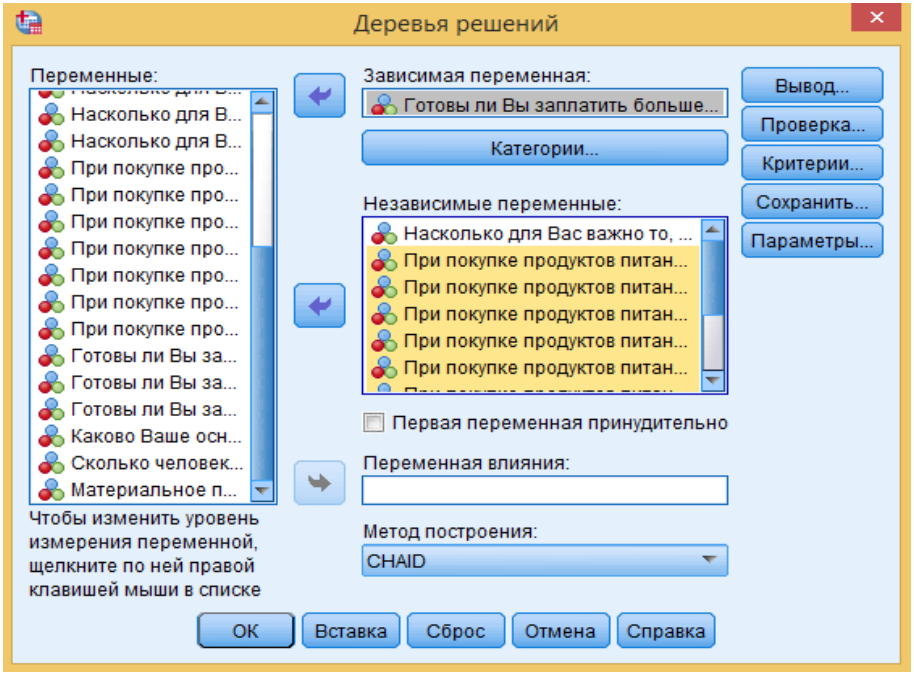
3. РЕАЛИЗАЦИЯ CHAID АНАЛИЗА В SPSS

3.1 ПРИМЕР CHAID В SPSS

3.1 Пример CHAID в SPSS

Задача: Выяснить, за какие характеристики продуктов питания потребители готовы платить больше?

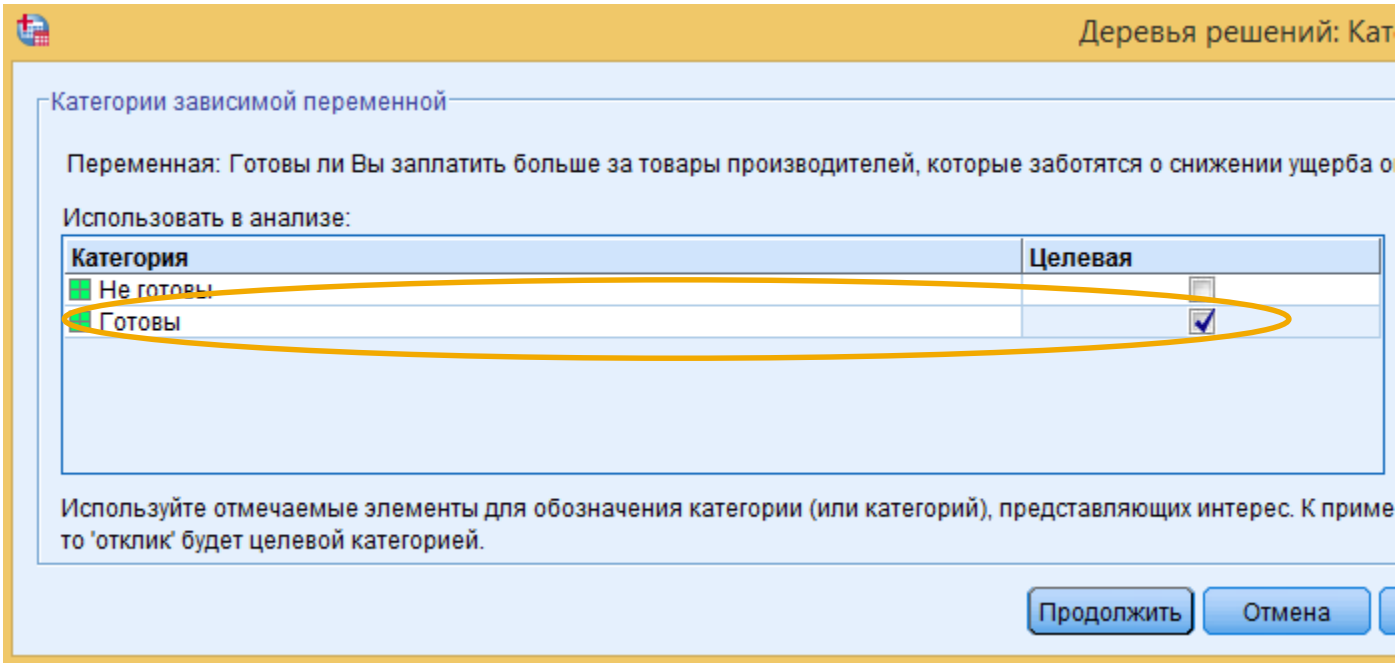
1. Открыть массив данных Ecology.sav.
2. Команды **«Анализ»** → **«Классификация»** → **«Деревья классификации».**
3. Зависимая переменная Q3C переносится в поле **«Группировать по».**
4. В качестве независимых переменных возьмем Q1C, Q2_1, Q2_4, Q2_5, Q2_6, Q2_8, Q2_9, Q2_10, Q2_12, Q2_14, Q2_15 и перенесем в поле **«Независимые переменные».**



3.1 Пример CHAID в SPSS

Следующим шагом необходимо выбрать, какая категория будет целевой. В нашем случае – это готовность платить больше за экологичные товары.

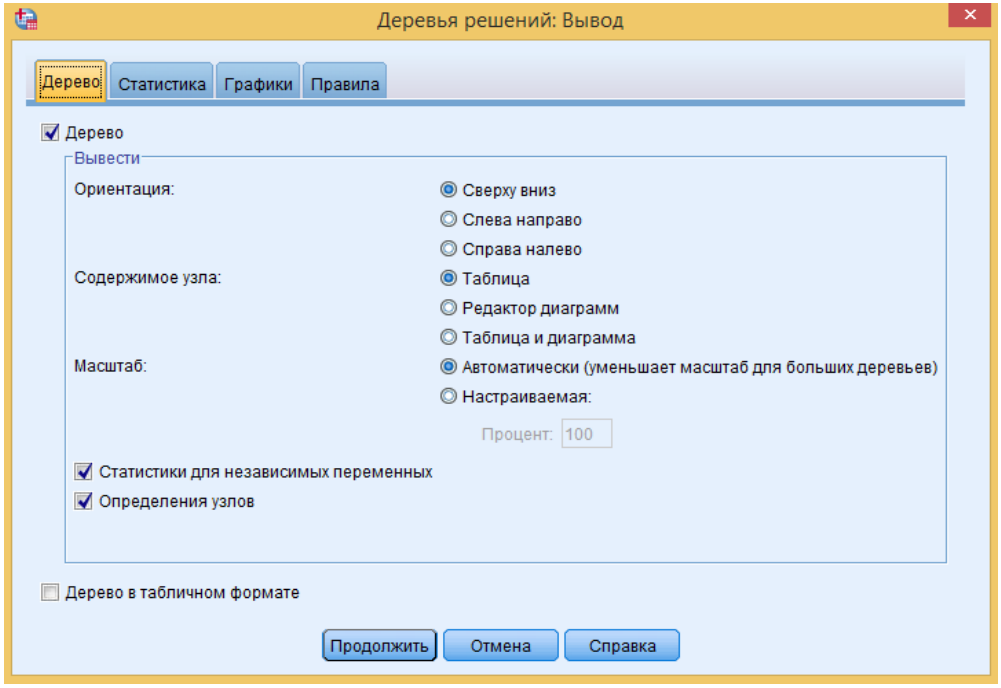
5. Щелкнуть по **«Категории»** и поставить галочку справа от категории **«Готовы».**



3.1 Пример CHAID в SPSS

6. Щелкнуть по кнопке **«Вывод».**

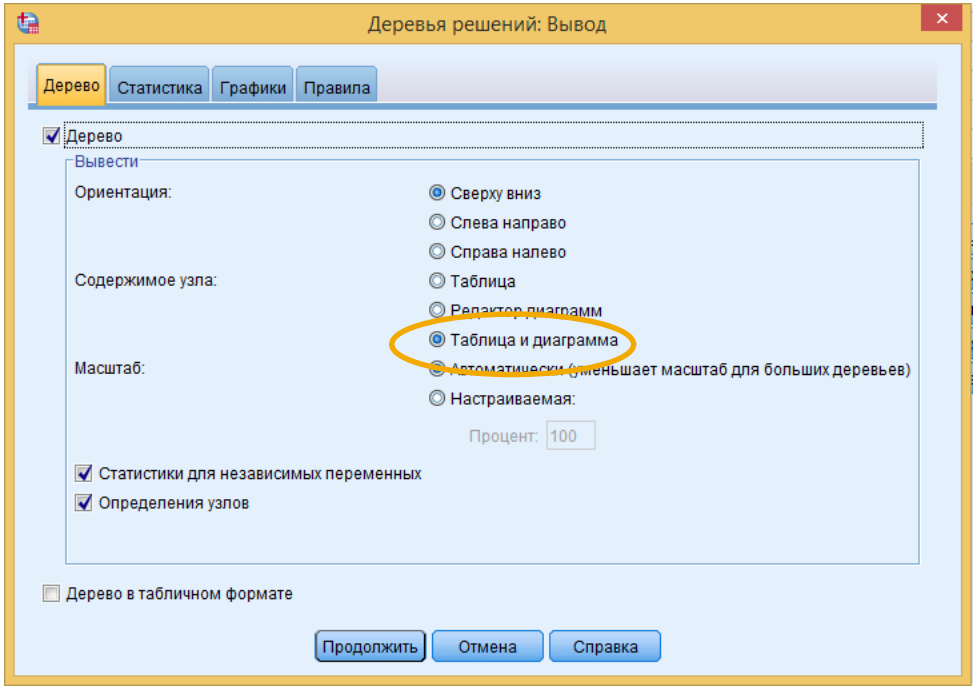
Здесь задается появление дерева решений и генерация таблиц. Можно запросить дополнительную статистическую информацию о модели, графическую интерпретацию соответствующих статистик, также можно запросить генерацию правил классификации для модели в SPSS синтаксисе, в SQL или в обычном текстовом формате.



3.1 Пример CHAID в SPSS

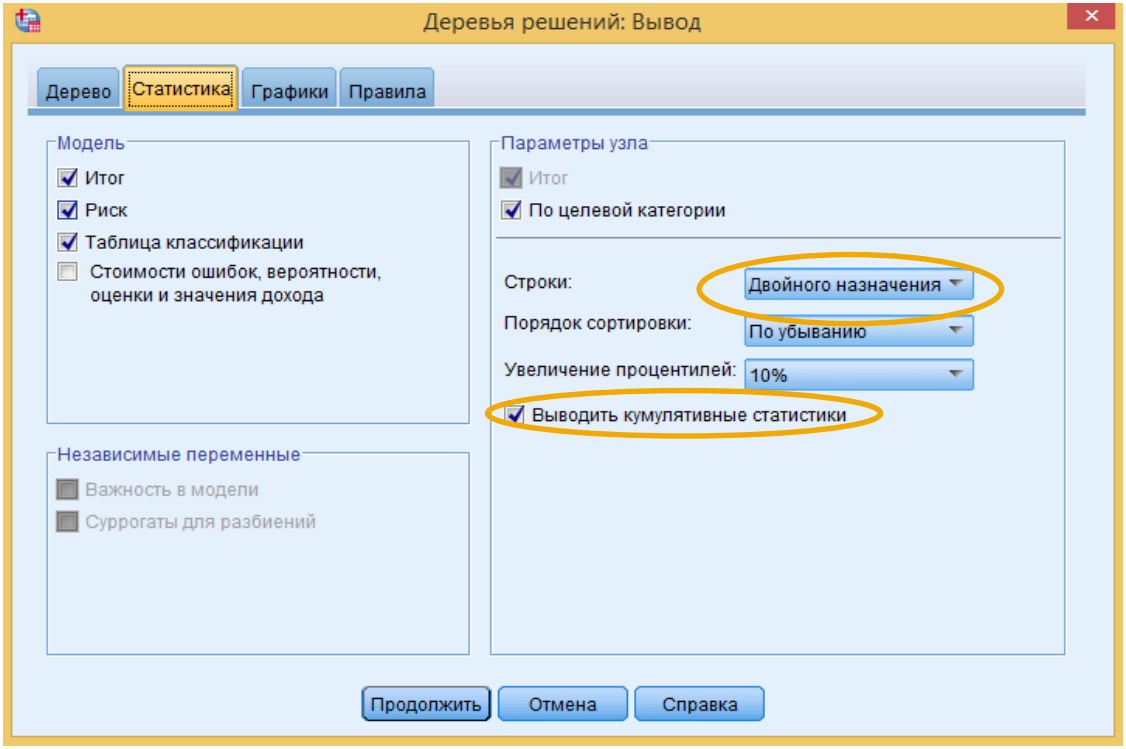
Можно выбрать в разделе **«Содержимое узла»** – **«Таблица и диаграмма»**, чтобы нагляднее представить предлагаемые статистики. Графическое изображение может быть полезным, так как можно увидеть, какие узлы лучше или хуже представляют целевую категорию без обращения к точным цифрам и статистикам.

Однако для большого дерева с множеством узлов такие диаграммы будут не читабельны.



3.1 Пример CHAID в SPSS

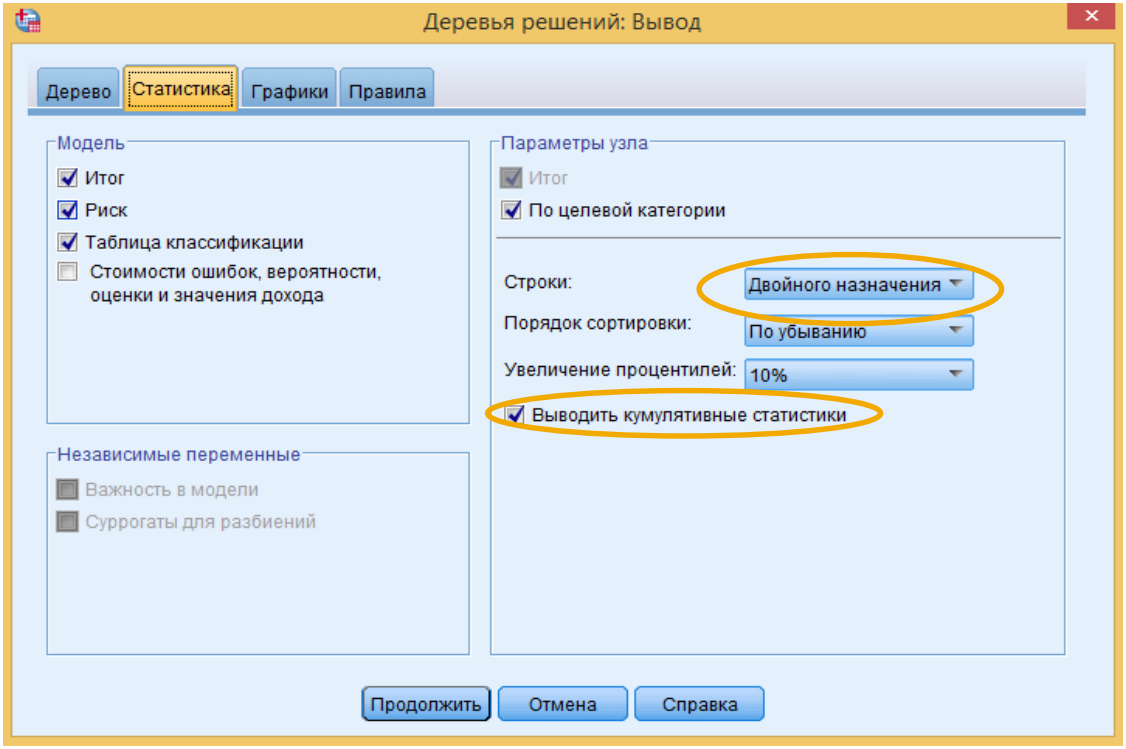
7. Перейти во вкладку **«Статистика».**
8. В параметре **«Строки»** изменить «Терминальные узлы» на **«Двойного назначения».**
9. Отметить **«Вывод кумулятивных статистик».**



3.1 Пример CHAID в SPSS

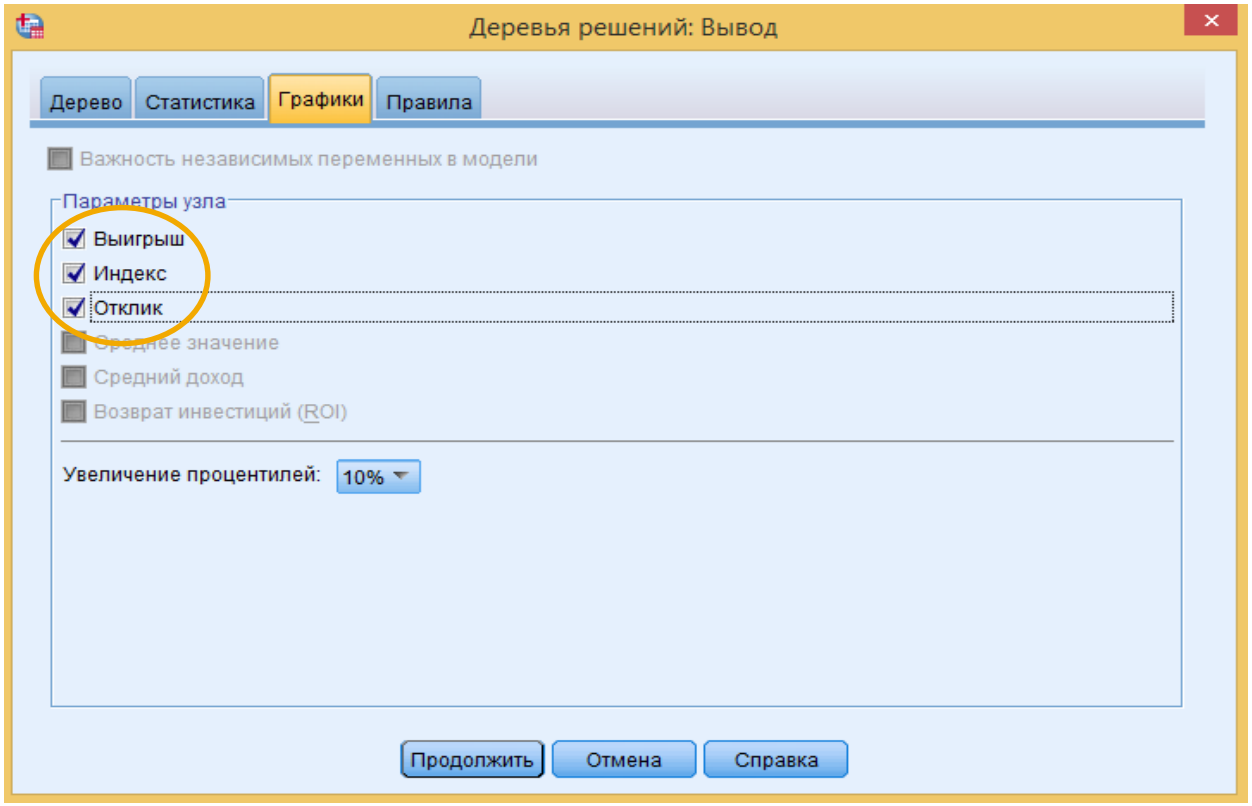
В разделе **«Параметры узла»** можно в **«Строки»** выбрать один из трех параметров: **«Терминальные узлы»**, **«Процентили»** и **«Двойного назначения»,** если хотим вывести терминальные узлы и процентили.

При порядке сортировки **«По убыванию»** терминальные узлы отсортируются так, что вверху таблицы окажутся узлы, наилучшим образом представляющие целевую категорию.



3.1 Пример CHAID в SPSS

10. Во вкладке **«Графики»** отметить все три доступных графических параметра.
11. В последней вкладке **«Правила»** оставить настройки по умолчанию.



3.1 Пример CHAID в SPSS

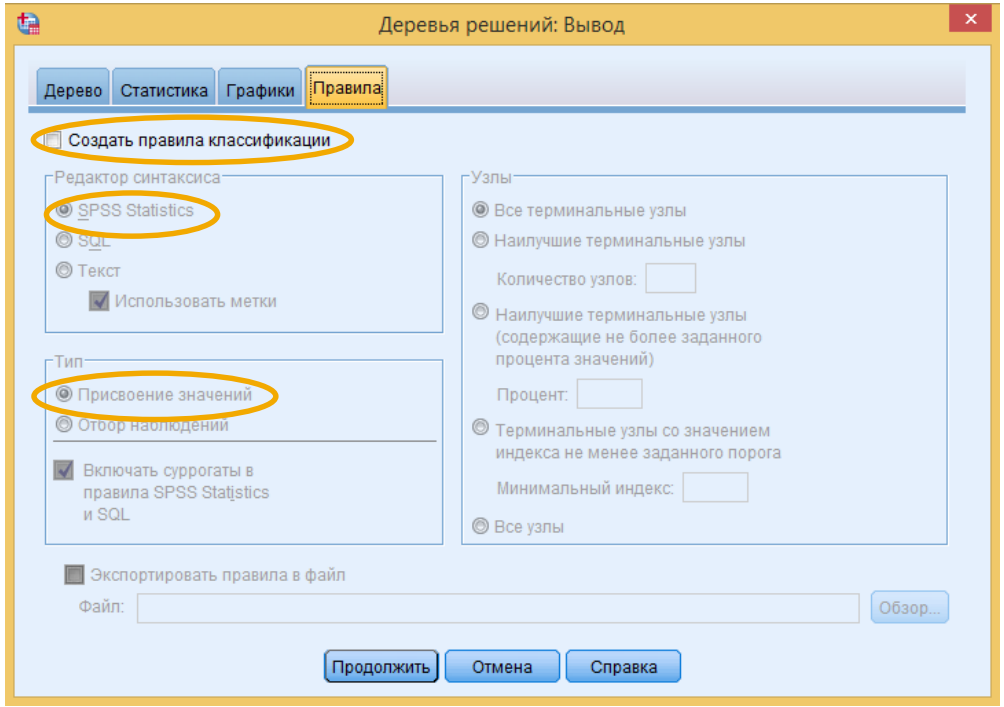
Во вкладке **«Правила»** можно сгенерировать правила отбора или классификации/ предсказания в виде командного синтаксиса SPSS, SQL или простого текста, чтобы в дальнейшем просмотреть их или экспортировать для применения к базам данных.

При выборе «SPSS Statistics» и типа «Присвоение значений» образуются новые переменные:

- nod_001 – номер узла, в который попадает наблюдение

- pre_001 – категория зависимой переменной, предсказанная для наблюдений данного узла

- prb_001 – доля наблюдений с предсказанной категорией в каждом узле.



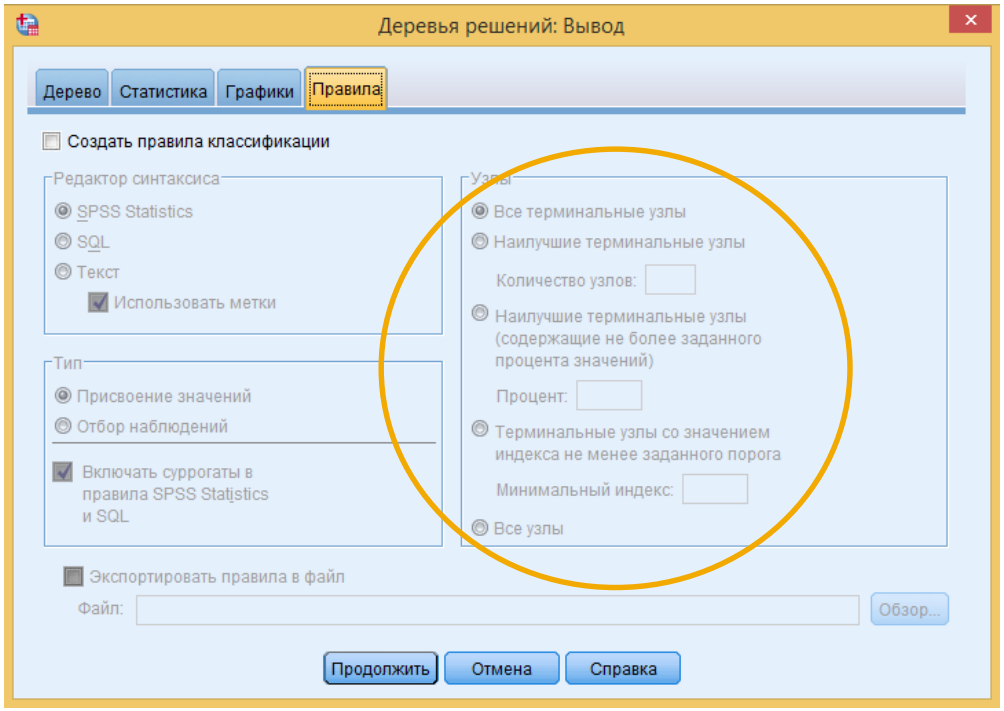
3.1 Пример CHAID в SPSS

Наилучшие терминальные узлы – можно выбрать N лучших по значению индекса термальных узлов.

Наилучшие термальные узлы (содержащие не более заданного процента значений) – то же самое, только вместо точного количества указывается процент наблюдений

Терминальные узлы со значением индекса не менее заданного порога – отбирает наилучшие узлы, для которых значение индекса больше или равно указанного минимального индекса

Все узлы – используется для присвоения значений, если нужно для дальнейшего анализа.



3.1 Пример CHAID в SPSS

12. Щелкнуть по кнопке **«Критерии» (**задает значения, которые используются в построении модели, такие как минимальное количество наблюдений в каждой группе или сегменте и уровень значимости, используемый в статистических тестах) и увеличить ограничения на наблюдения в узле до 500 и 250 (в виду большой выборки в рассматриваемом примере).

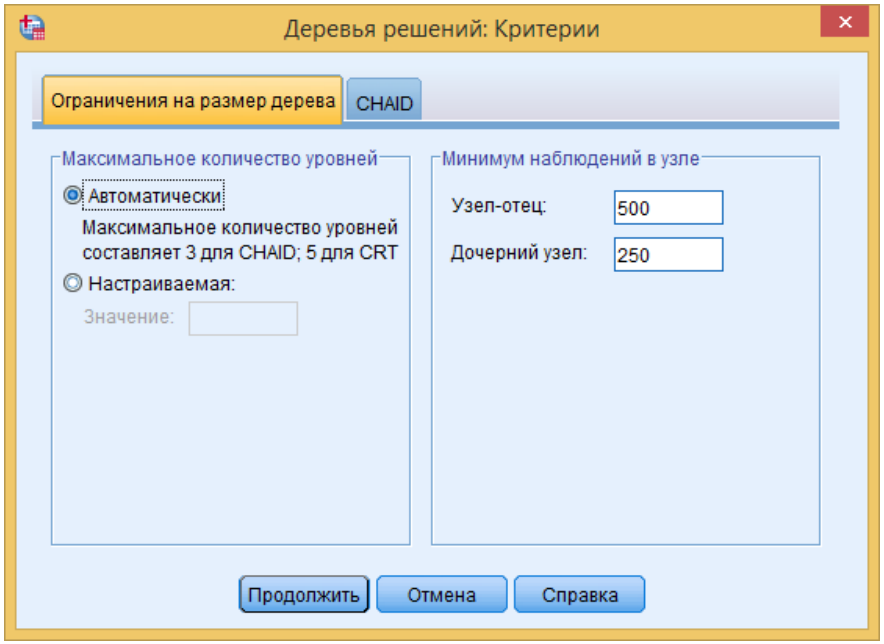
«Максимальное количество уровней» – количество слоев ниже корневого узла.

«Минимум наблюдений в узле»:

По умолчанию стоит 100 и 50

- если количество наблюдений в узле-отце менее 100, то дальнейший анализ не проводится;
- если в узле-сыне содержится меньше 50 наблюдений, то он не будет создаваться.

Данные критерии подходят для набора данных среднего размера.



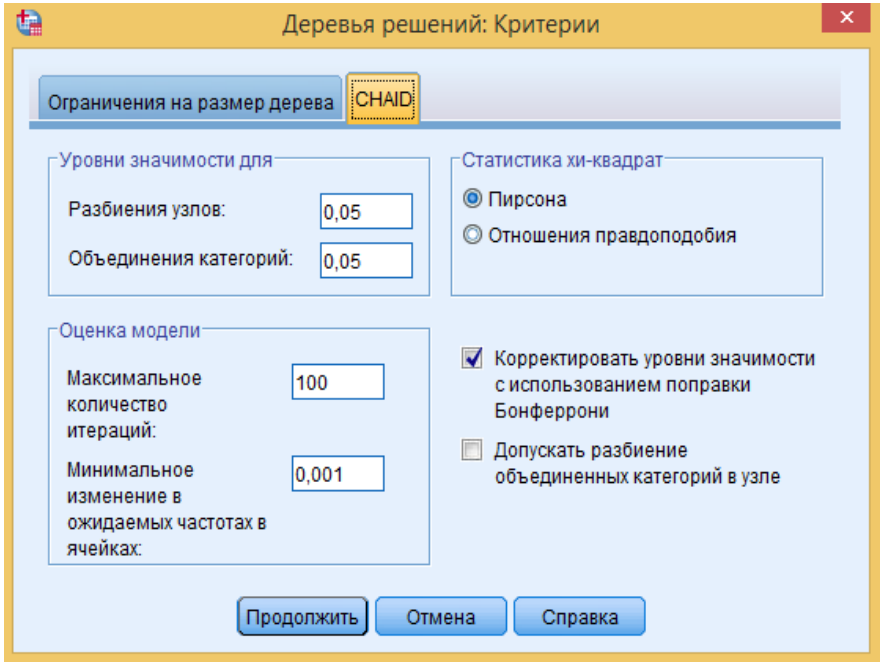
3.1 Пример CHAID в SPSS

13. Во вкладке **«CHAID»** оставить настройки по умолчанию.

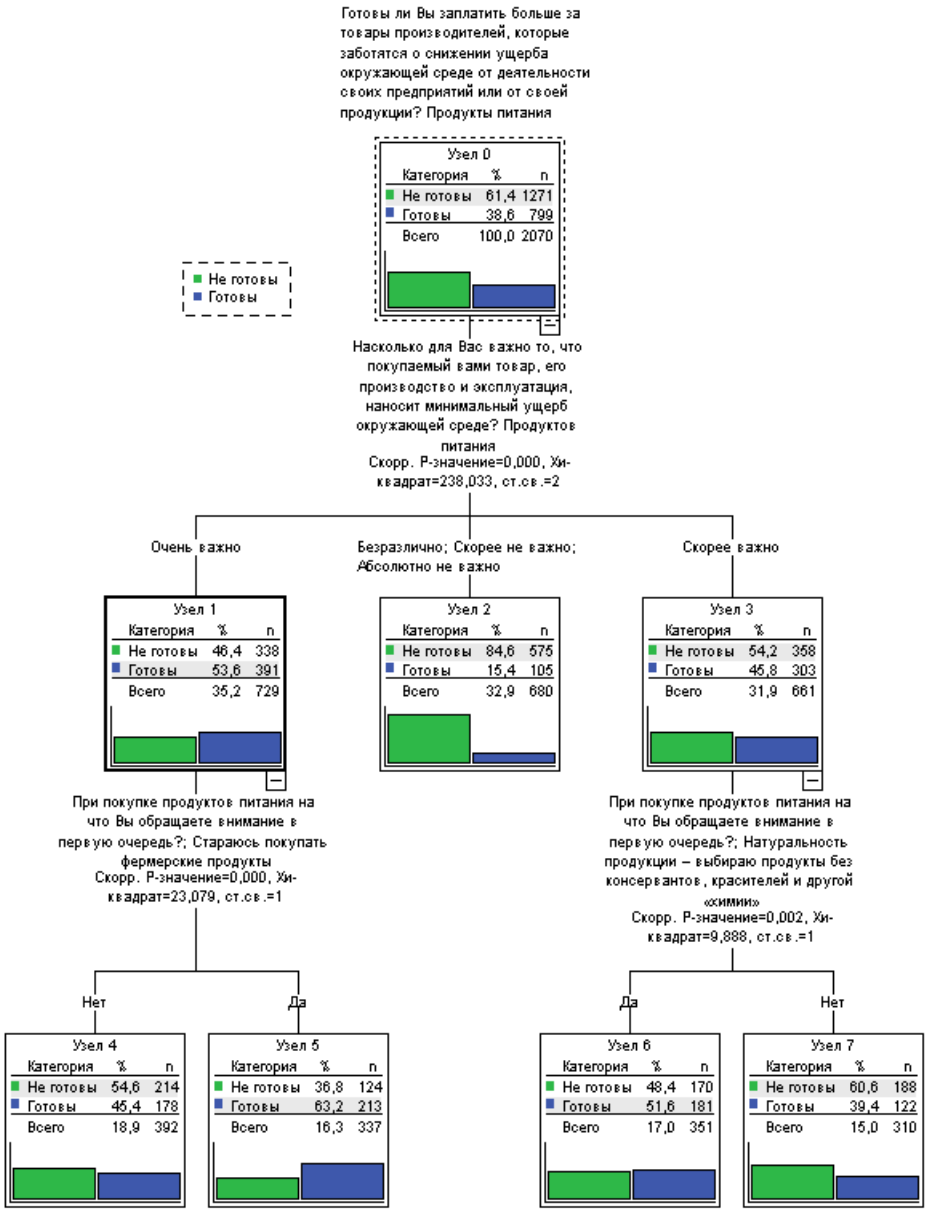
Уровень значимости по умолчанию составляет 0,05 для каждого теста хи-квадрат. Если нет значимых на указанном уровне отличий между группами, то они объединяются в одну категорию.

Хи-квадрат Пирсона – работает быстрее, но требует осторожности при работе с маленькими выборками

Отношение правдоподобия – более устойчивый, но требует более длительных вычислений. Предпочтительнее для маленьких выборок.



3.1 Пример CHAID в SPSS



3.1 Пример CHAID в SPSS

	Обозначения для таблиц
Узел	Номер узла – используется для нахождения узла на диаграмме дерева
Узел, N	Количество единиц наблюдения
Узел, %	Процент людей в данном узле от исходной выборки
Выигрыш, N	Количество единиц наблюдения в данном узле, которые попадают в целевую категорию
Выигрыш, %	Процент наблюдений в целевой категории для данного узла от общего объема целевой категории в выборке
Отклик	Процент наблюдений в целевой категории для данного узла от общего количества наблюдений в узле
Индекс	Отношение отклика в узле к отклику по выборке в целом

3.1 Пример CHAID в SPSS

Из таблицы видно, что узел 5 является наилучшим.

- В него попадает приблизительно **1/6** исходной выборки (337 из 2070) наблюдений.
- Из колонки Отклик видно, что **63,2%** наблюдений из узла 5 относятся к категории «готовы платить больше».
- Индекс показывает, что выбирая данный узел, ожидается найти в нем более чем в **1,5** раза больше наблюдений, относящихся к категории «готовы платить больше», чем в целом по выборке.

Выигрыши для узлов

Узел	По узлам						Суммарно					
	Узел		Выигрыш		Отклик	Индекс	Узел		Выигрыш		Отклик	Индекс
	N	Проценты	N	Проценты			N	Проценты	N	Проценты		
5	337	16,3%	213	26,7%	63,2%	163,7%	337	16,3%	213	26,7%	63,2%	163,7%
6	351	17,0%	181	22,7%	51,6%	133,6%	688	33,2%	394	49,3%	57,3%	148,4%
4	392	18,9%	178	22,3%	45,4%	117,6%	1080	52,2%	572	71,6%	53,0%	137,2%
7	310	15,0%	122	15,3%	39,4%	102,0%	1390	67,1%	694	86,9%	49,9%	129,4%
2	680	32,9%	105	13,1%	15,4%	40,0%	2070	100,0%	799	100,0%	38,6%	100,0%

Метод построения: CHAID

Зависимая переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

3.1 Пример CHAID в SPSS

- Если индекс больше 100%, то больше шансов выбрать наблюдение, попавшее в целевую категорию в данном узле, чем в выборке в целом.
- Если индекс меньше 100% – выбор в данном узле не даст преимущества по сравнению со случайным выбором из всей выборки.

- В «По узлам» выводится информация по каждому узлу, а в «Суммарно» – кумулятивные статистики по терминальным узлам.

Выигрыши для узлов

	По узлам						Суммарно					
	Узел		Выигрыш				Узел		Выигрыш			
Узел	N	Проценты	N	Проценты	Отклик	Индекс	N	Проценты	N	Проценты	Отклик	Индекс
5	337	16,3%	213	26,7%	63,2%	163,7%	337	16,3%	213	26,7%	63,2%	163,7%
6	351	17,0%	181	22,7%	51,6%	133,6%	688	33,2%	394	49,3%	57,3%	148,4%
4	392	18,9%	178	22,3%	45,4%	117,6%	1080	52,2%	572	71,6%	53,0%	137,2%
7	310	15,0%	122	15,3%	39,4%	102,0%	1390	67,1%	694	86,9%	49,9%	129,4%
2	680	32,9%	105	13,1%	15,4%	40,0%	2070	100,0%	799	100,0%	38,6%	100,0%

Метод построения: CHAID

Зависимая переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

3.1 Пример CHAID в SPSS

- Таблица **«Выигрыши для процентилей»** похожа на предыдущую таблицу **«Выигрыши для узлов»**, однако вместо характеристик для отдельных узлов выборка расщеплена на процентили.
- Например, можно заметить, что 30% всей выборки находится в 5 и 6 узлах. И ожидаемый отклик от 621 человека составит 57,9% (или 359 человек будут готовы заплатить больше).
- Они представляют 45% всех людей, которые бы были готовы заплатить больше. Этот результат в 1,5 раза превышает отклик по сравнению со случайным отбором 30% людей.

Выигрыши для процентилей

Процентиль	Узлы	N	Выигрыш		Отклик	Индекс
			N	Проценты		
10	5	207	131	16,4%	63,2%	163,7%
20	5 ; 6	414	253	31,6%	61,0%	158,1%
30	6	621	359	45,0%	57,9%	150,0%
40	6 ; 4	828	458	57,3%	55,3%	143,2%
50	4	1035	552	69,0%	53,3%	138,1%
60	4 ; 7	1242	636	79,6%	51,2%	132,6%
70	7 ; 2	1449	703	88,0%	48,5%	125,7%
80	2	1656	735	92,0%	44,4%	115,0%
90	2	1863	767	96,0%	41,2%	106,7%
100	2	2070	799	100,0%	38,6%	100,0%

Метод построения: CHAID
Зависимая переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

3.1 Пример CHAID в SPSS

- В таблице **«Риск»** выводятся общие значения ошибок классификации. Это является важной оценкой прогностической точности дерева.

- Также в таблице выводится значение среднеквадратической ошибки оценки классификации, с помощью которого можно рассчитать доверительные интервалы, умножив среднеквадратическую ошибку на 1,96 для 95% доверительных интервалов и прибавив/вычесть это значение к значению ошибки классификации.

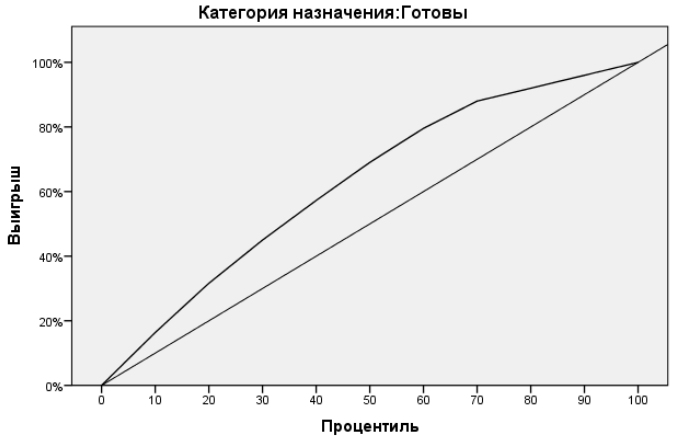
Ошибка классификации находилась бы в пределах **от 31,84% до 35,76%**

Риск

Оценка	Среднеквадратичная ошибка
,338	,010

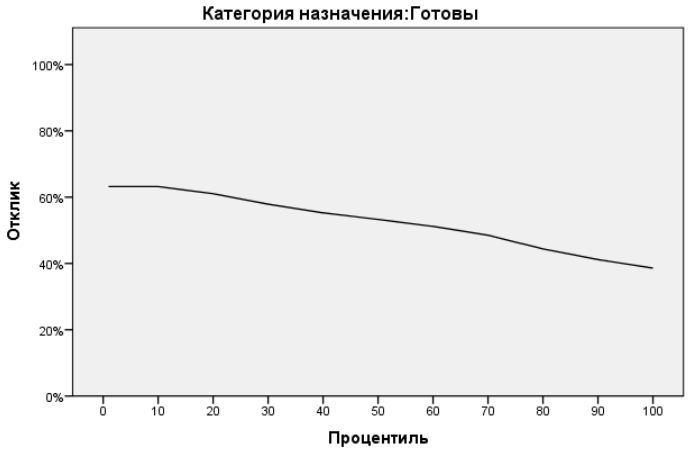
Метод построения: CHAID
Зависимая переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

3.1 Пример CHAID в SPSS



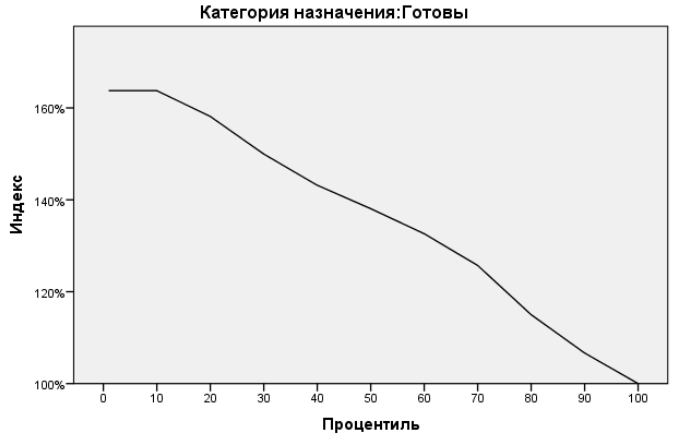
Метод построения:CHAID

Зависимая переменная:Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания



Метод построения:CHAID

Зависимая переменная:Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания



Метод построения:CHAID

Зависимая переменная:Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

- На каждой диаграмме отображаются в графическом виде результаты из соответствующего столбца «Выигрыш для процентилей».
- Диагональная линия на диаграмме «Выигрыш» - наблюдения при случайном отборе.

3.1 Пример CHAID в SPSS

- **«Таблица классификации»** является таблицей сопряженности, в которой указывается число правильно и неправильно классифицированных наблюдений.
- В каждом узле берется категория с наибольшим числом откликов и рассчитывается, сколько наблюдений было правильно отнесено к данной доминирующей категории

- Общее число корректно классифицируемых наблюдений равно 977+394 = 1371, что составляет **66,2%.**

Классификация

Наблюденные	Предсказанные		
	Не готовы	Готовы	Процент правильных
Не готовы	977	294	76,9%
Готовы	405	394	49,3%
Общая процентная доля	66,8%	33,2%	66,2%

Метод построения: CHAID
Зависимая переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

3.2 ПРОВЕРКА АДЕКВАТНОСТИ МОДЕЛИ

3.2 Проверка адекватности модели

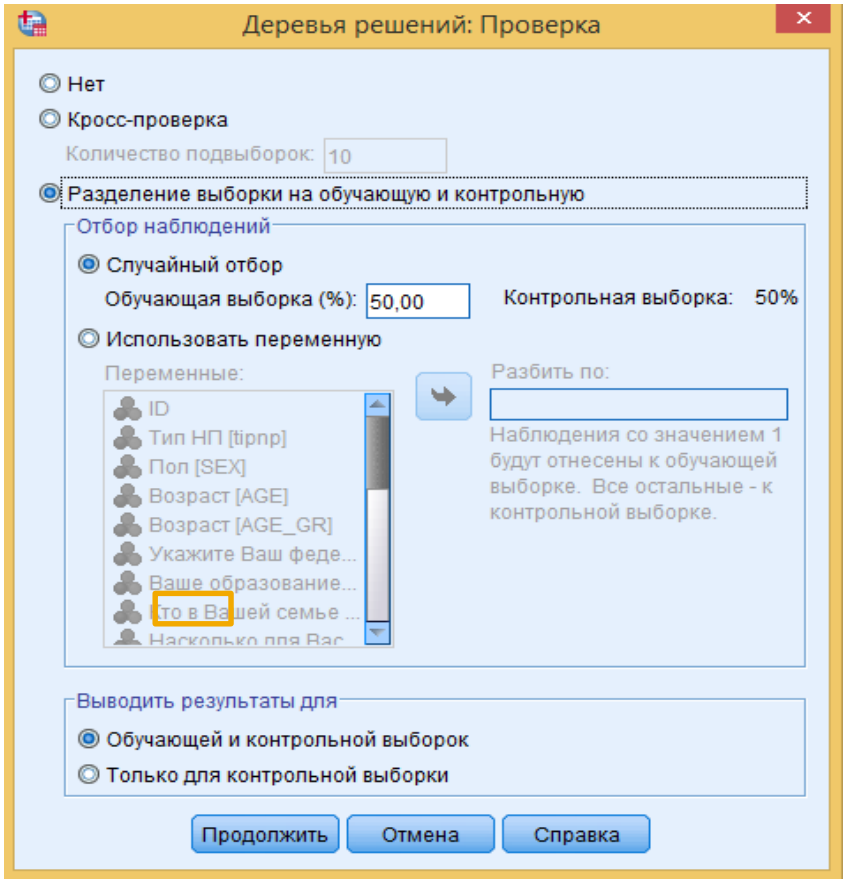
При анализе методом **CHAID** существует две опции для проверки модели:

- **Разбиение** – разбивка данных на две части: одна для построения модели, а вторая – для проверки (применяется при большой выборке).

- **N-кратная перекрестная проверка** – применяется в случае небольшой выборки (менее 1000 наблюдений). Весь набор данных разбивается на n-выборок (обычно 10) приблизительно равного объема. Затем строятся n-деревьев с последовательным исключением каждой из выборок.

3.2 Проверка адекватности модели

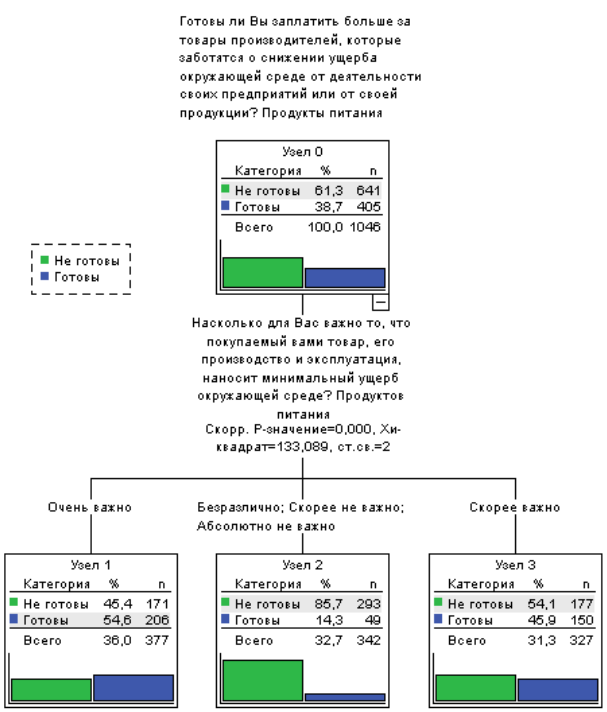
1. Команды **«Анализ» → «Классификация» → «Деревья классификации».**
2. **«Вывод» → «Строки»** изменить на **«Терминальные узлы».**
3. Щелкнуть по кнопке **«Проверка»** и выбрать **«Разделение выборки на обучающую и контрольную».**
4. Нажать Продолжить
5. Нажать ОК



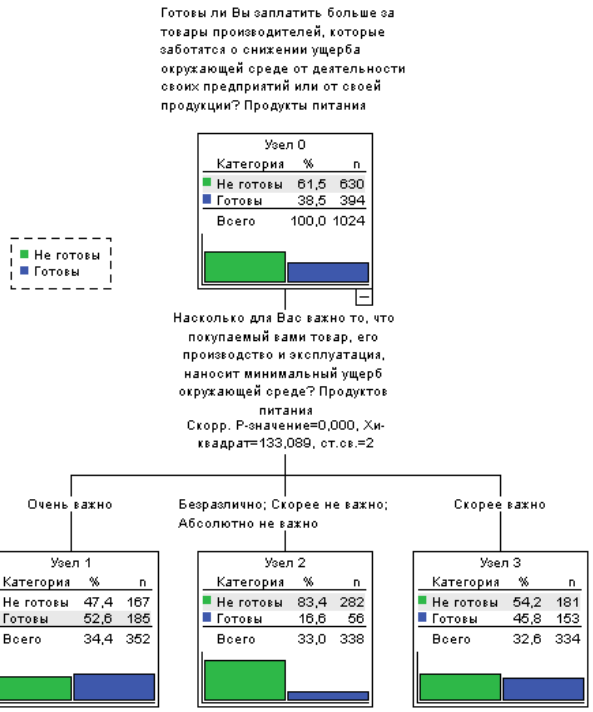
3.2 Проверка адекватности модели

Мы получили две **Диаграммы дерева** – отдельно для обучающей и контрольной выборок. Отметим, что они практически идентичны за исключением небольших различий в количестве наблюдений в узлах. Однако, полученные диаграммы отличаются от первоначального Дерева решений (полученного без использования проверки): новая модель включает только три терминальных узла, в то время как в старой модели их было пять.

Обучающая выборка



Контрольная выборка



3.2 Проверка адекватности модели

- Таблица **«Выигрыши для узлов»** также содержит результаты для обучающей и контрольной выборок.
- Поскольку результаты в обеих выборках не сильно различаются, то сделаем вывод о том, что построенная модель является достаточно общей.

Если узлы с наилучшими результатами по обучающей выборке существенно отличаются по доле отклика, значит, модель не прошла проверку и будет плохо работать на новых выборках.

Выигрыши для узлов

		По узлам						Суммарно					
		Узел		Выигрыш				Узел		Выигрыш			
Пример	Узел	N	Проценты	N	Проценты	Отклик	Индекс	N	Проценты	N	Проценты	Отклик	Индекс
Обучение	1	377	36,0%	206	50,9%	54,6%	141,1%	377	36,0%	206	50,9%	54,6%	141,1%
	3	327	31,3%	150	37,0%	45,9%	118,5%	704	67,3%	356	87,9%	50,6%	130,6%
	2	342	32,7%	49	12,1%	14,3%	37,0%	1046	100,0%	405	100,0%	38,7%	100,0%
Критерий	1	352	34,4%	185	47,0%	52,6%	136,6%	352	34,4%	185	47,0%	52,6%	136,6%
	3	334	32,6%	153	38,8%	45,8%	119,1%	686	67,0%	338	85,8%	49,3%	128,1%
	2	338	33,0%	56	14,2%	16,6%	43,1%	1024	100,0%	394	100,0%	38,5%	100,0%

Метод построения: CHAID

Зависимая переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

3.2 Проверка адекватности модели

- Оценка риска для обучающей выборки составляет 0,354. Это значение незначительно отличается от оценки риска для контрольной выборки – 0,367.
- Отметим, что большие различия могут указывать на то, что дерево является недостаточно общим и устойчивым.

Ошибка классификации для обучающей выборки находилась бы в пределах **от 32,45% до 38,34%**

Ошибка классификации для контрольной выборки находилась бы в пределах **от 33,76% до 39,64%**

Риск

Пример	Оценка	Среднеквадратичная ошибка
Обучение	,354	,015
Критерий	,367	,015

Метод построения: CHAID
Зависимая переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

3.2 Проверка адекватности модели

- Результаты таблицы **«Классификация»** говорят нам о том, что потребители не готовы платить больше за какие-либо характеристики питания.
- В первоначальном варианте мы получили практически такой же результат - большинство респондентов были против.

Классификация

Пример	Наблюденные	Предсказанные		
		Не готовы	Готовы	Процент правильных
Обучение	Не готовы	470	171	73,3%
	Готовы	199	206	50,9%
	Общая процентная доля	64,0%	36,0%	64,6%
Критерий	Не готовы	463	167	73,5%
	Готовы	209	185	47,0%
	Общая процентная доля	65,6%	34,4%	63,3%

Метод построения: CHAID

Зависимая переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Продукты питания

Литература по Теме 11

1. **Курс Сегментация рынка в IBM SPSS Statistics. – М., 2014**
   - Глава 9. Анализ с помощью CHAID
   - Глава 10. Обобщения и дополнительные возможности CHAID