

**ТЕМА 11**

# Анализ с помощью CHAID

**Лабораторная работа 21 и 22**

# Лабораторная работа №21

## **Анализ с помощью CHAID**

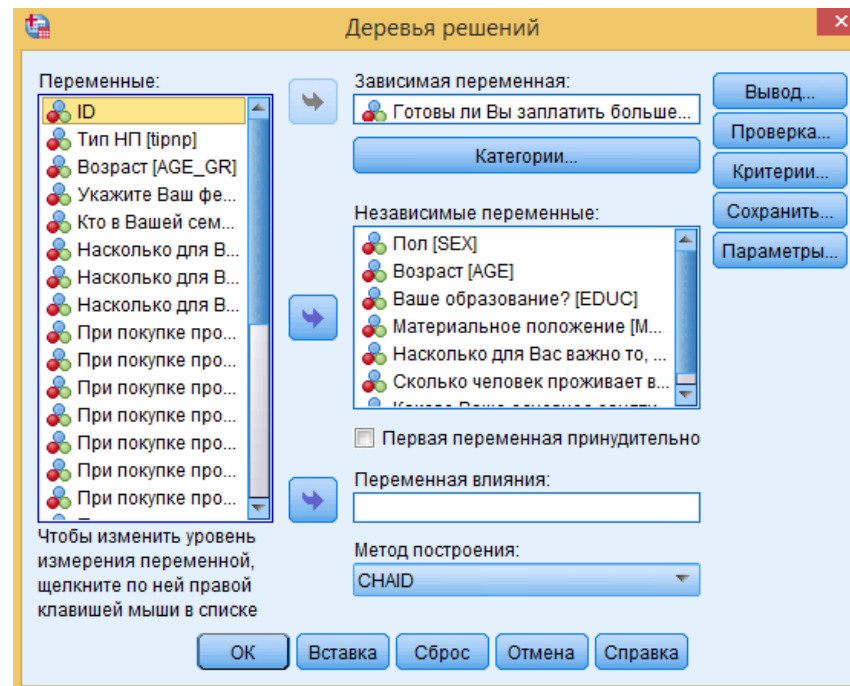
## Лабораторная работа №21

---

1. Откройте файл **Ecology.sav**.
2. Выясните, готовы ли потребители платить больше за экологичное производство одежды и обуви (переменная Q3\_A)?
3. Интерпретируйте полученные результаты.

## Лабораторная работа №21

1. Откройте массив данных Ecology.sav.
2. Команды **«Анализ»** → **«Классификация»** → **«Деревья классификации»**.
3. Зависимая переменная Q3A переносится в поле **«Группировать по»**.
4. В качестве независимых переменных возьмем SEX, AGE, EDUC, MAT\_POL, Q1A, FAMILY и перенесем в поле **«Независимые переменные»**.



## Лабораторная работа №21

Следующим шагом необходимо выбрать, какая категория будет целевой. В нашем случае – это важность экологичного производства.

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

Деревья решений: Кат

Категории зависимой переменной

Переменная: Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба о

Использовать в анализе:

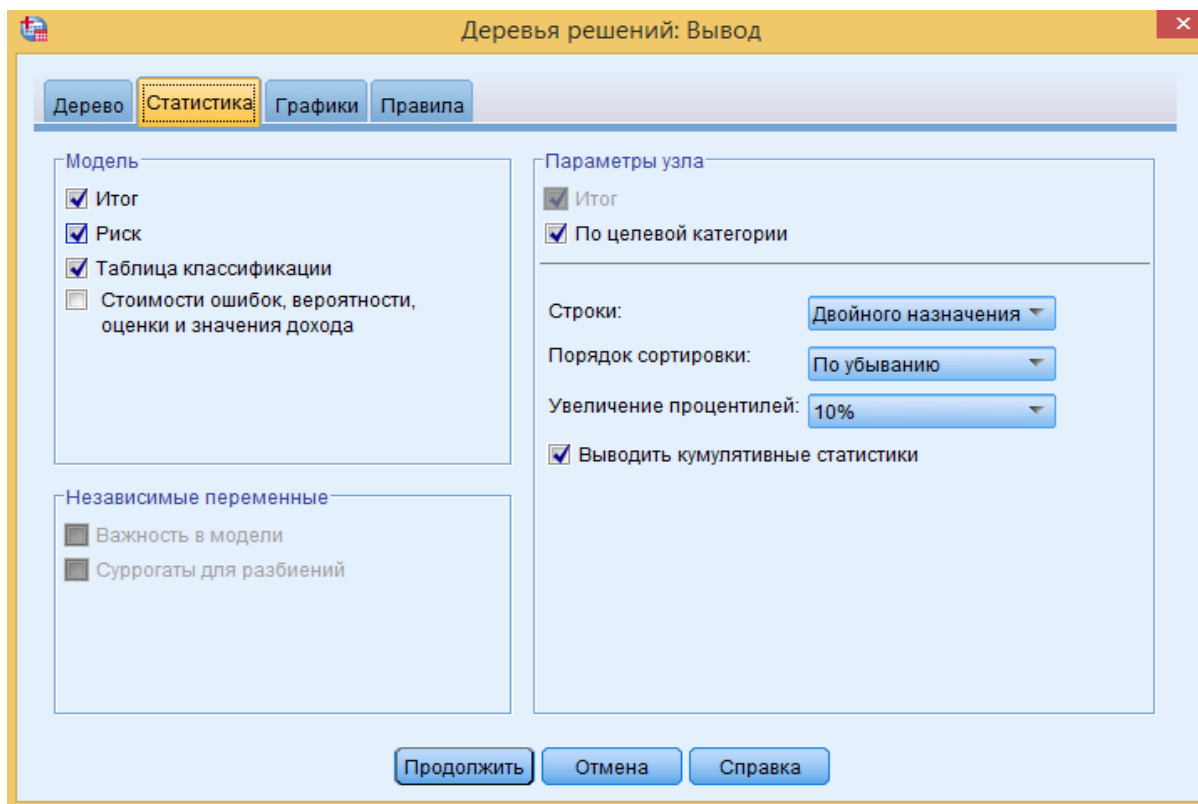
Категория	Целевая
<input checked="" type="checkbox"/> Не готовы	<input type="checkbox"/>
<input checked="" type="checkbox"/> Готовы	<input checked="" type="checkbox"/>

Используйте отмечаемые элементы для обозначения категории (или категорий), представляющих интерес. К приме то 'отклик' будет целевой категорией.

Продолжить Отмена

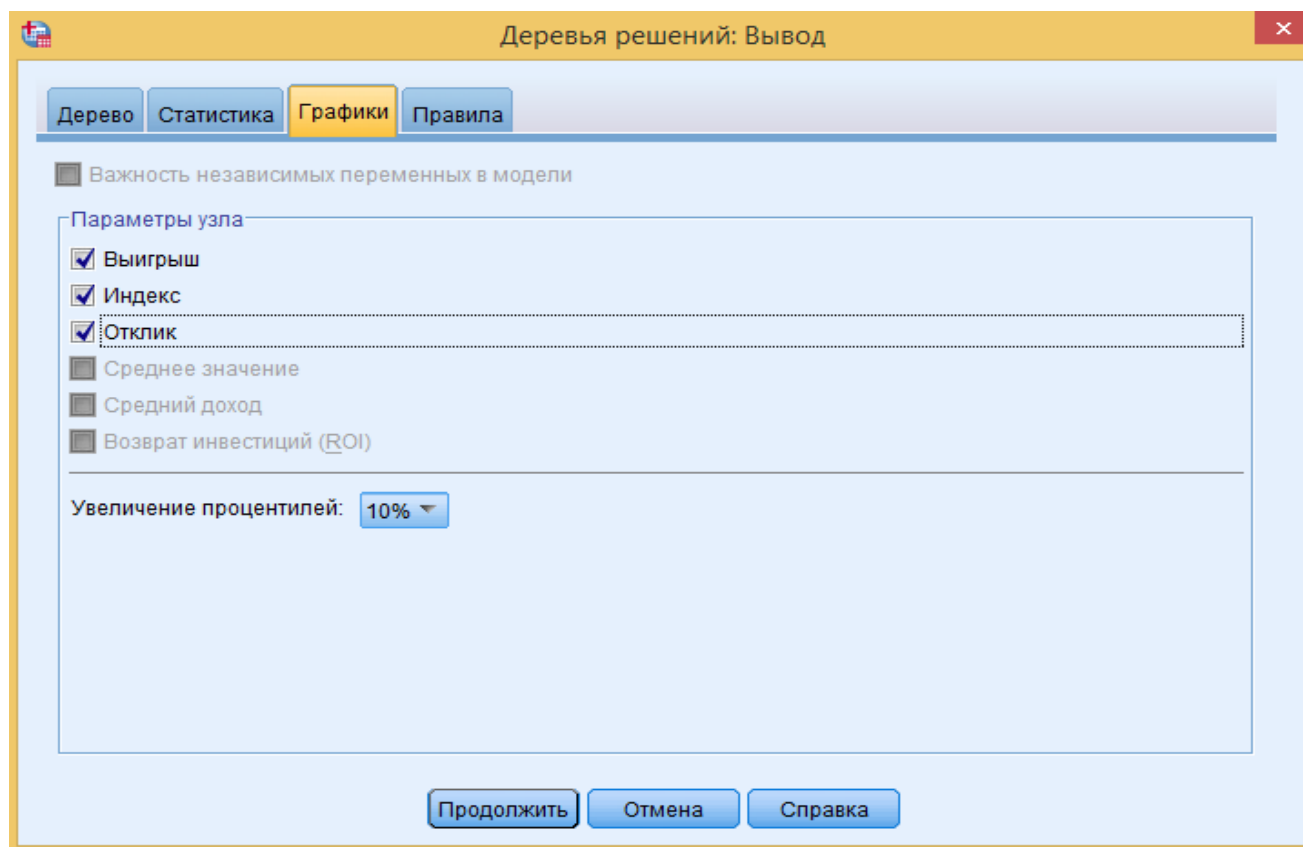
## Лабораторная работа №21

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



## Лабораторная работа №21

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



## Лабораторная работа №21

- Щелкнуть по кнопке «**Критерии**» и увеличить ограничения на наблюдения в узле до 500 и 250 (в виду большой выборки в рассматриваемом примере).
- Во вкладке «**CHAID**» оставить настройки по умолчанию.

Древья решений: Критерии

Ограничения на размер дерева CHAID

Максимальное количество уровней

Автоматически  
Максимальное количество уровней составляет 3 для CHAID; 5 для CRT

Настраиваемая:  
Значение:

Минимум наблюдений в узле

Узел-отец:

Дочерний узел:

Продолжить Отмена Справка

Древья решений: Критерии

Ограничения на размер дерева CHAID

Уровни значимости для

Разбиения узлов:

Объединения категорий:

Статистика хи-квадрат

Пирсона

Отношения правдоподобия

Оценка модели

Максимальное количество итераций:

Минимальное изменение в ожидаемых частотах в ячейках:

Корректировать уровни значимости с использованием поправки Бонферрони

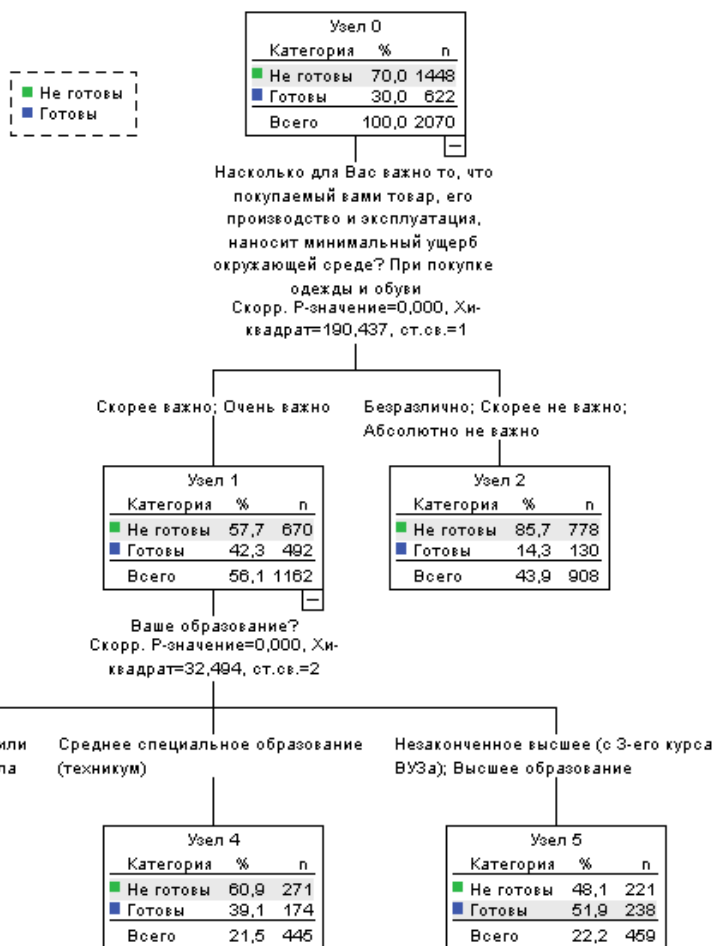
Допускать разбиение объединенных категорий в узле

Продолжить Отмена Справка



# Лабораторная работа №21

Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Одежда и обувь



- Таким образом, мы выяснили, что 70% потребителей не готовы платить больше за экологичное производство.
- При этом важность нанесения минимального вреда окружающей среде при производстве одежды и обуви имеет значение для 56% потребителей.
- Среди тех, кто имеет высшее и незаконченное высшее образование, 51,9% готовы платить больше.

## Лабораторная работа №21

Выигрыши для узлов

Узел	По узлам						Суммарно					
	Узел		Выигрыш		Отклик	Индекс	Узел		Выигрыш		Отклик	Индекс
	N	Проценты	N	Проценты			N	Проценты	N	Проценты		
5	459	22,2%	238	38,3%	51,9%	172,6%	459	22,2%	238	38,3%	51,9%	172,6%
4	445	21,5%	174	28,0%	39,1%	130,1%	904	43,7%	412	66,2%	45,6%	151,7%
3	258	12,5%	80	12,9%	31,0%	103,2%	1162	56,1%	492	79,1%	42,3%	140,9%
2	908	43,9%	130	20,9%	14,3%	47,6%	2070	100,0%	622	100,0%	30,0%	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? Одежда и обувь

Выигрыши для процентилей

Процентиль	Узлы	N	Выигрыш		Отклик	Индекс
			N	Проценты		
10	5	207	107	17,3%	51,9%	172,6%
20	5	414	215	34,5%	51,9%	172,6%
30	5 ; 4	621	301	48,4%	48,5%	161,5%
40	4	828	382	61,5%	46,2%	153,7%
50	4 ; 3	1035	453	72,8%	43,7%	145,5%
60	3 ; 2	1242	503	80,9%	40,5%	134,9%
70	2	1449	533	85,7%	36,8%	122,4%
80	2	1656	563	90,5%	34,0%	113,1%
90	2	1863	592	95,2%	31,8%	105,8%
100	2	2070	622	100,0%	30,0%	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? Одежда и обувь

Потребители, которые готовы платить больше за экологичное производство одежды, обуви, находятся в пятом узле, которые не готовы – во втором

## Лабораторная работа №21

- На уровне доверительной вероятности 95% ошибка классификации находится в пределах **от 27,24% до 31,16%**.
- Общее число корректно классифицируемых наблюдений равно  $1227+238 = 1465$ , что составляет **70,8%**.

**Риск**

Оценка	Среднеквадратичная ошибка
,292	,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? Одежда и обувь

**Классификация**

Наблюдаемые	Предсказанные		
	Не готовы	Готовы	Процент правильных
Не готовы	1227	221	84,7%
Готовы	384	238	38,3%
Общая процентная доля	77,8%	22,2%	70,8%

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? Одежда и обувь

# Лабораторная работа №22

## **Анализ с помощью CHAID**

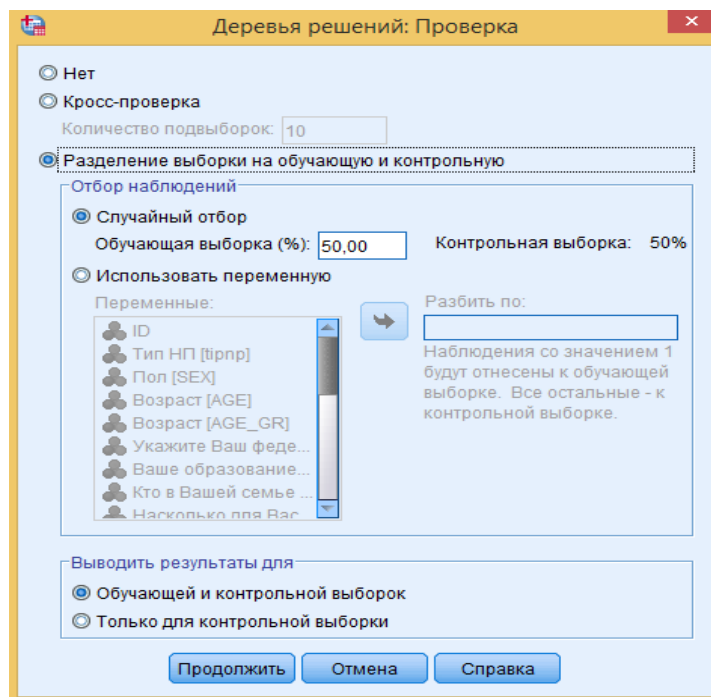
## Лабораторная работа №22

---

1. Откройте файл **Ecology.sav**.
2. Сделайте проверку адекватности модели по результатам Лабораторной работы №21.
3. Интерпретируйте полученные результаты.

## Лабораторная работа №22

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».



## Лабораторная работа №22

---

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»**.

## Обучающая выборка

Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Одежда и обувь

■ Не готовы  
■ Готовы

Узел 0		
Категория	%	n
■ Не готовы	68,9	737
■ Готовы	31,1	333
Всего	100,0	1070

Насколько для Вас важно то, что покупаемый вами товар, его производство и эксплуатация, наносит минимальный ущерб окружающей среде? При покупке одежды и обуви  
Скорр. Р-значение=0,000, Хи-квадрат=117,172, ст.св.=1

Скорее важно; Очень важно

Узел 1		
Категория	%	n
■ Не готовы	55,6	339
■ Готовы	44,4	271
Всего	57,0	610

Безразлично; Скорее не важно; Абсолютно не важно

Узел 2		
Категория	%	n
■ Не готовы	86,5	398
■ Готовы	13,5	62
Всего	43,0	460

## Контрольная выборка

Готовы ли Вы заплатить больше за товары производителей, которые заботятся о снижении ущерба окружающей среде от деятельности своих предприятий или от своей продукции? Одежда и обувь

■ Не готовы  
■ Готовы

Узел 0		
Категория	%	n
■ Не готовы	71,1	711
■ Готовы	28,9	289
Всего	100,0	1000

Насколько для Вас важно то, что покупаемый вами товар, его производство и эксплуатация, наносит минимальный ущерб окружающей среде? При покупке одежды и обуви  
Скорр. Р-значение=0,000, Хи-квадрат=117,172, ст.св.=1

Скорее важно; Очень важно

Узел 1		
Категория	%	n
■ Не готовы	60,0	331
■ Готовы	40,0	221
Всего	55,2	552

Безразлично; Скорее не важно; Абсолютно не важно

Узел 2		
Категория	%	n
■ Не готовы	84,8	380
■ Готовы	15,2	68
Всего	44,8	448



## Лабораторная работа №22

В результате мы имеем 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 2 узла, в то время как в старой модели их было 5.

- Поскольку результаты в обеих выборках не сильно различаются, то сделаем вывод о том, что построенная модель является достаточно общей.

Выигрыши для узлов

Пример	Узел	Узел		Выигрыш		Отклик	Индекс
		N	Проценты	N	Проценты		
Обучение	1	610	57,0%	271	81,4%	44,4%	142,8%
	2	460	43,0%	62	18,6%	13,5%	43,3%
Критерий	1	552	55,2%	221	76,5%	40,0%	138,5%
	2	448	44,8%	68	23,5%	15,2%	52,5%

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? Одежда и обувь

## Лабораторная работа №22

---

- Оценка риска для обучающей выборки составляет 0,311. Это значение незначительно отличается от оценки риска для контрольной выборки – 0,289.
- На уровне доверительной вероятности 95% ошибка классификации для обучающей выборки находится в пределах **от 28,36% до 33,84%**, а для контрольной выборки – в пределах **от 26,16% до 31,64%**

**Риск**

Пример	Оценка	Среднеквадратичная ошибка
Обучение	,311	,014
Критерий	,289	,014

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? Одежда и обувь