

## XVIII. Дискриминантный анализ

- [Пример из области медицины](#)
- [Пример из области социологии](#)
- [Пример из области биологии](#)
- [Пример из области биологии \(три группы\)](#)

С помощью дискриминантного анализа на основании некоторых признаков (независимых переменных) индивидуум может быть причислен к одной из двух (или к одной из нескольких) заданных заранее групп.

Такая постановка задачи, в особенности в случае двух заранее заданных групп, очень сильно напоминает постановку задачи для метода логистической регрессии (см. гл. 16.4). Ядром дискриминантного анализа является построение так называемой дискриминантной функции

$$d = b_1x_1 + b_2x_2 + \dots + b_nx_n + a,$$

где  $x_1$  и  $x_n$  — значения переменных, соответствующих рассматриваемым случаям, константы  $b_1$ - $b_n$  и  $a$  — коэффициенты, которые и предстоит оценить с помощью дискриминантного анализа. Целью является определение таких коэффициентов, чтобы по значениям дискриминантной функции можно было с максимальной четкостью провести разделение по группам.

### 18.1. Пример из области медицины

Обратимся ещё раз к примеру, который уже приводился при рассмотрении логистической регрессии. В этом примере приводятся выборочные данные о пациентах с нарушениями работы легких. Эти данные хранятся в файле `lunge.sav`. Приведем ещё раз переменные, которые в данном случае будут применяться при дискриминантном анализе:

Имя переменной	Значение
out	Исход (0 = скончался, 1 = выжил)
alter	Возраст
bzeit	Время проведения искусственного дыхания в часах
kob	Концентрация кислорода в смеси для искусственного дыхания
адд	Интенсивность искусственного дыхания
geschl	Пол (1 = мужской, 2 = женский)
gr	Рост

Переменная `out` делит пациентов на две группы; при помощи остальных переменных предстоит прогнозировать принадлежность к одной из групп.

- Откройте файл `lunge.sav`.
- Выберите в меню `Analyze (Анализ) Classify (Классифицировать) Discriminant...` (Дискриминантный анализ)

Откроется диалоговое окно `Discriminant Analysis (Дискриминантный анализ)`.

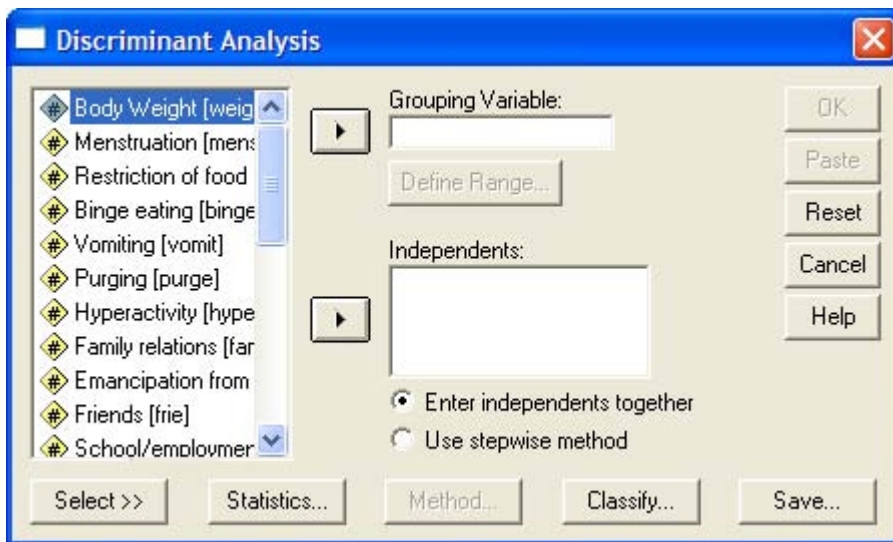


Рис. 18.1: Диалоговое окно *Discriminant Analysis* (Дискриминантный анализ).

- Поместите переменную out в поле, предназначенное для групповых переменных.
- После щелчка по выключателю Define Range... (Определить промежуток) введите минимальное и максимальные значения этой переменной: 0 и 1.
- Переменным agg, alter, bzeit, gcschl, gr и kob присвойте статус независимых переменных. Для начала оставим установленный по умолчанию метод: Enter independents together (Одновременный учет всех независимых переменных), при котором в анализе одновременно будут участвовать все независимые переменные.
- После щелчка по выключателю Statistics... (Статистики) активируйте опции: Means (Средние значения), Univariate ANOVAs (Одномерные тесты ANOVA), Unstandardized Function Coefficients (Нестандартизированные коэффициенты функции) и Within-group Correlation Matrice (Корреляционная матрица внутри группы).
- Через выключатель Classify (Классифицировать) сделайте дополнительно запрос на вывод диаграмм по отдельным группам (Separate-groups Plots), результатов для отдельных наблюдений (Casewise results) и сводной таблицы (Summary table). При выводе результатов для отдельных наблюдений ограничимся первыми двадцатью, поместив этот предел в соответствующую позицию диалогового окна.

Довольно полезный график для объединенных групп, который был реализован в ранних версиях SPSS, и сейчас можно активировать в диалоговом окне, однако вместо графика в окне отображения результатов будет появляться предупреждение о том, что такая гистограмма в анализах более не доступна.

- При помощи выключателя Save... (Сохранить) активируйте сохранение значения дискриминантной функции в дополнительной переменной (Discriminant Scores).
- Начните расчёт нажатием ОК.

После вводного обзора действительных и пропущенных значений приводятся средние значения, стандартные отклонения, количество наблюдений для каждой группы в отдельности и суммарные показатели для обеих групп.

Переменная geschl является при этом дихотомической переменной, принадлежащей к номинальной шкале с кодировками: 1 (мужской пол) и 2 (женский пол). Средние значения для обеих групп по исходу Легения, кажущиеся на первый взгляд бесполезными, равны 1,63492 и 1,45588; если бы вместо этого переменные были закодированы при помощи 0 и 1, то оба средних значения равнялись бы 0,63492 и 0,45588 соответственно. Для таких дихотомических переменных, закодированных при помощи 0 и 1, среднее значение указывает на долю наблюдений с кодировкой 1. Это означает, что для группы "скончался" доля женщин в процентном отношении составляет 63,492, а для группы "выжил" 45,588.

<b>Group Statistics (Статистики для групп)</b>					
<b>Outcome (Исход)</b>		<b>Mean (Среднее значение)</b>	<b>Std. Deviation (Стандартное отклонение)</b>	<b>Valid N (listwise) (Действительные значения (по списку))</b>	
				<b>Unweighted (Не взвешено)</b>	<b>Weighted (Взвешено)</b>
gesto-rben (Скончался)	Aggressivitaet der Beatmung (Интенсивность искусственного дыхания)	15,90013	10,90013	63	63,000
	ALTER (Возраст)	31,92063	13,82529	63	63,000
	Beatmungszeit in Std. (Время проведения искусственного дыхания в часах)	15,36508	10,50085	63	63,000
	Geschlecht (Пол)	1,63492	,48532	63	63,000
	Koerper-groesse (Рост)	165,1429	15,55931	63	63,000
	Sauerstoff-Konzentration (Концентрация кислорода в смеси для искусственного дыхания)	,85952	,14807	63	63,000
ueberlebt (Выжил)	Aggressivitaet der Beatmung (Интенсивность искусственного дыхания)	11,69699	8,16057	68	68,000
	ALTER (Возраст)	27,97059	10,86411	68	68,000
	Beatmungszeit in Std. (Время проведения искусственного дыхания в часах)	10,79412	5,10065	68	68,000
	Geschlecht (Пол)	1,45588	,50175	68	68,000
	Koerper-groesse (Рост)	172,0588	11,01137	68	68,000
	Sauerstoff-Konzentration (Концентрация кислорода в смеси для искусственного дыхания)	,80338	,15493	68	68,000
Total	Aggressivitaet der Beatmung (Интенсивность искусственного дыхания)	13,51843	9,72600	131	131,000
	ALTER (Возраст)	29,87023	12,48654	131	131,000
	Beatmungszeit in Std. (Время проведения искусственного дыхания в часах)	12,99237	8,44120	131	131,000
	Geschlecht (Пол)	1,54198	,50015	131	131,000
	Koerper-groesse (Рост)	168,7328	13,78339	131	131,000
	Sauerstoff-Konzentration (Концентрация кислорода в смеси для искусственного дыхания)	,83038	,15369	131	131,000

Затем проводится тест, насколько значимо различаются между собой переменные в обеих группах; наряду с тестовой величиной, в качестве которой служит Лямбда Уилкса ("Wilks-Lambda"), применяется также и простой дисперсионный анализ. Для всех переменных (кроме возраста, для которого однако также просматривается сильная тенденция к значимости) получается значимое различие между обеими группами:

Tests of Equality of Group Means (Тест равенства групповых средних значений)

	Wilks Lambda (Лямбда Уилкса)	F	df1	df2	Sig. (Знач.)
Aggressivitaet der Beatmung (Интенсивность искусственного дыхания)	,962	5,116	1	129	,025
ALTER (Возраст)	,975	3,331	1	129	,070
Beatmungszeit in Std. (Время проведения искусственного дыхания в часах)	,926	10,273	1	129	,002
Geschlecht (Пол)	,968	4,297	1	129	,040
Koerpergroesse (Рост)	,937	8,722	1	129	,004
Sauerstoff-Konzentration (Концентрация кислорода в смеси для искусственного дыхания)	,966	4,481	1	129	,036

Далее следует корреляционная матрица между всеми переменными, в которой приводятся коэффициенты, осредненные для обеих групп:

Pooled Within-Groups Matrices (Объединённые внутригрупповые матрицы)

	Aggressivitaet der Beatmung (Интенсивность искусств. дыхания)	ALTER (Возраст)	Beatmungszeit in Std. (Время провед. искусств. дыхания в часах)	Geschlecht (Пол)	Koerpergroesse (Рост)	Sauerstoff-Konzentrat (Концентрат кислорода в смеси для искусственного дыхания)	
Correlation (Корреляция)	Aggressivitaet der Beatmung (Интенсивность искусственного дыхания)	1,000	-,072	-,058	,141	-,042	,285
	ALTER (Возраст)	-,072	1,000	,093	-,040	,277	-,119
	Beatmungszeit in Std. (Время проведения искусственного дыхания в часах)	-,058	,093	1,000	,069	-,126	-,089
	Geschlecht (Пол)	,141	-,040	,069	1,000	-,481	-,066
	Koerpergroesse (Рост)	-,042	,277	-,126	-,481	1,000	,000
	Sauerstoff-Konzentration (Концентрация кислорода в смеси для искусственного дыхания)	,285	-,119	-,089	-,066	,000	1,000

Следующими шагами являются расчёт и анализ коэффициентов дискриминантной функции. Значения этой функции должны как можно отчетливее разделять обе группы. Мерой удачности

этого разделения служит корреляционный коэффициент между рассчитанными значениями дискриминантной функции и показателем принадлежности к группе:

### Eigenvalues (Собственные значения)

Function (Функция)	Eigenvalue (Собственное значение)	% of Variance (% дисперсии)	Cumulative % (Совокупный %)	Canonical Correlation (Каноническая корреляция)
1	,256"	100,0	100,0	,452

a. First 1 canonical discriminant functions were used in the analysis (В этом анализе используются первые 1 канонические дискриминантные функции).

Wilks' Lambda (Лямбда Уилкса)

Test of Function(s) (Тест функции (и))	Wilks' Lambda (Лямбда Уилкса)	Chi-square (Хи-квадрат)	df	Sig. (Значимость)
1	,796	28,733	6	,000

Судя по значению коэффициента, равному 0,452, корреляция абсолютно не удовлетворительная. При помощи Лямбда Уилкса производится тест на то, значимо ли в обеих группах отличаются друг от друга средние значения дискриминантной функции; в приводимом примере, значение  $p < 0,001$ , указывает на очень значимое различие.

Значение, выводимое под именем "Eigenvalue" (Собственное значение), соответствует отношению суммы квадратов между группами к сумме квадратов внутри групп. Эти две суммы Вы сможете получить, если проведете дисперсионный анализ значений дискриминантной функции (переменная dis1\_1) по фактору out (см. гл. 13.3). Большие собственные значения (в данном случае такого, к сожалению, не наблюдается) указывают на "хорошие" (удачно подобранные) дискриминантные функции.

Следующая таблица дает представление о том, как сильно отдельные переменные, применяемые в дискриминантной функции, коррелируют со стандартизованными значениями этой дискриминантной функции. При этом корреляционные коэффициенты были рассчитаны в обеих группах по отдельности и затем усреднены:

### Standardized Canonical Discriminant Function Coefficients

(Стандартизированные канонические коэффициенты дискриминантной функции)	
	Function (Функция)
	1
Aggressivitaet der Beatmung (Интенсивность искусственного дыхания)	,316
ALTER (Возраст)	,494
Beatmungszeit in Std. (Время проведения искусственного дыхания в часах)	,491
Geschlecht (Пол)	,066
Koerpergroesse (Рост)	-,544
Sauerstoff-Konzentration (Концентрация кислорода в смеси для искусственного дыхания)	,385

## Structure Matrix

<b>(Структурная матрица)</b>	
	<b>Function (Функция)</b>
	1
Beatmungszeit in Std. (Время проведения искусственного дыхания в часах)	,558
Koerpergroesse (Рост)	-,514
Aggressivitaet der Beatmung (Интенсивность искусственного дыхания)	,393
Sauerstoff-Konzentration (Концентрация кислорода в смеси для искусственного дыхания)	,368
Geschlecht (Пол)	,361
ALTER (Возраст)	,318

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions (Объединённые корреляции внутри групп между дискриминантными переменными и стандартизированными каноническими дискриминантными функциями).

Variables ordered by absolute size of correlation within function (Переменные расположены в соответствии с абсолютными корреляционными величинами внутри функции).

И в заключение, приводятся сами коэффициенты дискриминантной функции:

### Canonical Discriminant Function Coefficients

<b>(Канонические коэффициенты дискриминантной функции)</b>	
	<b>Function (Функция)</b>
	1
Aggressivitaet der Beatmung (Интенсивность искусственного дыхания)	,033
ALTER (Возраст)	,040
Beatmungszeit in Std. (Время проведения искусственного дыхания в часах)	,060
Geschlecht (Пол)	,133
Koerpergroesse (Рост)	-,041
Sauerstoff-Konzentration (Концентрация кислорода в смеси для искусственного дыхания)	2,539
(Constant)	2,121

### Unstandardized coefficients (Нестандартизированные коэффициенты)

Здесь речь идёт о нестандартизированных коэффициентах — это множители при заданных значениях переменных, входящих в дискриминантную функцию. Стандартизированные коэффициенты, которые приводились ранее, основаны на стандартизированных значениях переменных, получаемых с помощью z-преобразования.

Далее приводятся средние значения дискриминантной функции в обеих группах:

## Functions at Group Centroids

(Функции групповых центроидов)	
Outcome (Исход)	Function (функция)
	1
gestorben (Скончался)	,522
ueberlebt (Выжил)	-,483

Unstandardized canonical discriminant functions evaluated at group means (Нестандартизированные канонические дискриминантные функции, которые оцениваются по групповым средним значениям).

Далее следует таблица, в которой построчно для каждого наблюдения приводится информация о значении дискриминантной функции и определяется принадлежность к одной из двух групп. Мы здесь ограничились первыми двадцатью наблюдениями.

Группа, к которой фактически принадлежит наблюдение, отображается в колонке с именем "Actual Group" (Фактическая группа). В следующих трёх колонках содержится информация о прогнозе принадлежности к группе, сделанном на основании значения дискриминантной функции. Сначала приводится прогнозируемая принадлежность к группе; если она не соответствует фактической принадлежности, то в колонке "Predicted

\*\* Misclassified case (Неправильно классифицированное наблюдение);

Далее выводятся две вероятности. Вторая из этих двух вероятностей, обозначенная  $P(G=g|D=d)$ , является мерой принадлежности к одной из двух групп. Это вероятность того, что некоторым наблюдением принадлежит к прогнозированной группе, которая рассчитывается на основе подстановки в дискриминантную функцию значений набора переменных, соответствующих данному наблюдению. Вероятность того, что данное наблюдение принадлежит к другой группе получается вычитанием меры принадлежности из 1. Она приводится в колонке с названием "Second Highest Group" (Вторая по старшинству группа). Если мы рассмотрим первое наблюдение, то здесь вероятность того, что данный пациент выживет, рассчитанная на основании значения исходных переменных, равна 0,702 (в действительности он скончался).

Первую из двух рассмотренных вероятностей, получившую название  $P(D>d|G=g)$ , называют ещё и условной вероятностью. Это вероятность того, что пациент, принадлежащий к прогнозируемой группе, действительно имеет значения параметров, соответствующие дискриминантной функции или некоторые другие крайние значения.

В другой колонке приводится квадрат расстояния Махаланобиса до центроида (среднего значения группы значений дискриминантной функции). В правой колонке таблицы приводится соответствующее значение дискриминантной функции. Распределение значения дискриминантной функции отдельно по группам изображается на двух отдельных гистограммах.

Можно заметить, что значения дискриминантной функции для первой группы (скончался) смещены вправо, а значения второй группы (выжил) — влево, что однако свидетельствует об очень сильном смешении.

В завершении приводится классификационная таблица с указанием достигнутой точности прогнозирования. Значение этой точности равно 68,7 %, что является неудовлетворительным:

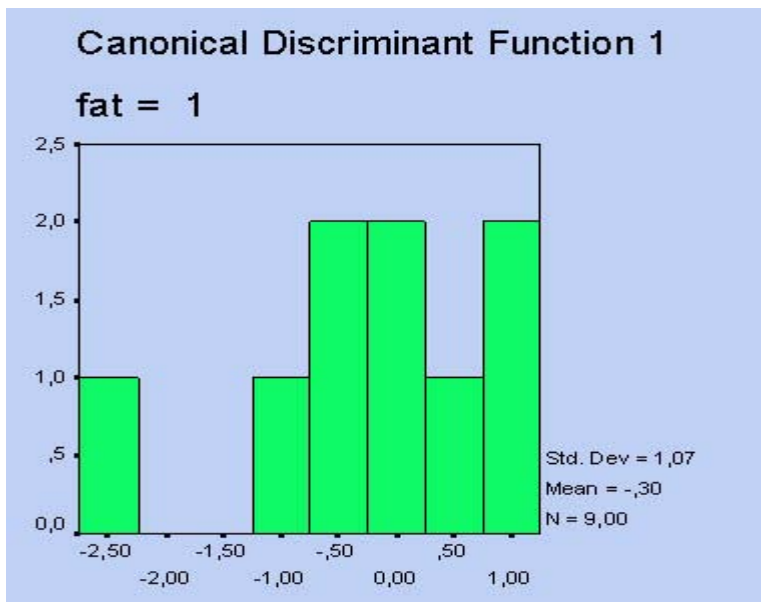


Рис. 18.2: Распределение значений дискриминантной функции для группы «скончался»

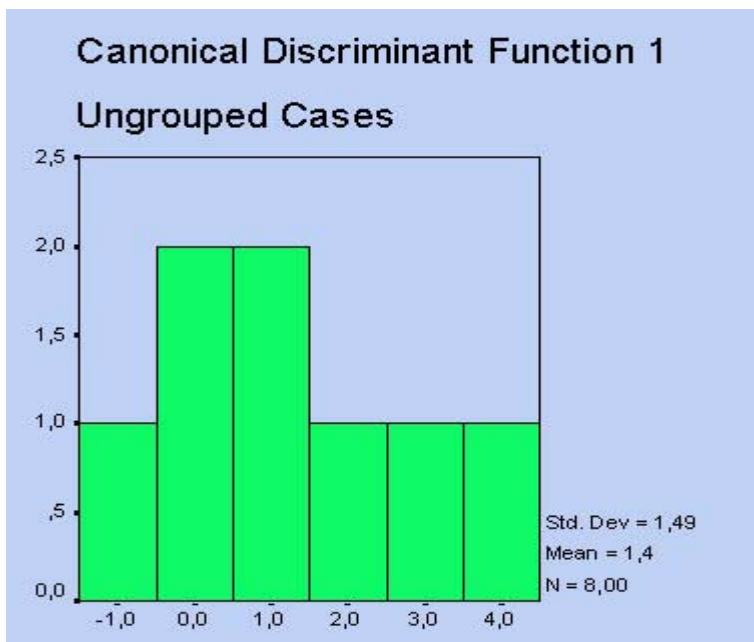


Рис. 18.3: Распределение значений дискриминантной функции для группы «выжил»

### Classification Results 3

(Классификационные результаты)					
		Outcome (Исход)	Predicted Group Membership (Предсказанная принадлежность к одной из групп)		Total (Сумма)
			gestorben (Скончался)	ueberlebt (Выжил)	
Original Первоначально	Count (Количество)	gestorben (скончался)	38	25	63
		ueberlebt (Выжил)	16	52	68



	%	gestorben (сконча-лся)	60,3	39,7	100,0
		ueberlebt (Выжил)	23,5	76,5	100,0

а. 68,7% of original grouped cases correctly classified (68,7 % первоначально сгруппированных наблюдений были классифицированы корректно).

При применении метода логарифмической регрессии (см. гл. 16.4) результат получился немного лучше (доля корректного прогноза 70,99 %).

Для случая, когда пациенту мужского пола, 25 лет, ростом 184 см искусственное дыхание делали на протяжении 5 часов, при концентрации кислорода равной 0,7 и интенсивности соответствующей значению 10, получается следующее значение дискриминантной функции

$$d = 2,121 + 0,033*10 + 0,04*25 + 0,06*5 + 0,133*1 - 0,041*184 + 2,539*0,7 = -1,883$$

Опираясь на распределение значений дискриминантной функции, этого пациента можно отнести к группе выживших.

При выполнении дискриминантного анализа, как и для других многомерных процедур, можно применять и пошаговый образ действий, который как раз и рекомендуется при наличии большого количества независимых переменных. Этот метод похож на многомерный регрессионный анализ, однако переменные при проведении дискриминантного анализа выбираются по другим критериям.

Рассчитаем ещё раз наш пример, но уже с применением пошагового метода.

- В исходном диалоговом окне дискриминантного анализа активируйте опцию Use stepwise method (Использовать пошаговый метод).
- Щёлкните на кнопке Method... (Метод)

Откроется диалоговое окно Discriminant Analysis: Step-wise Method (Дискриминантaый анализ: Пошаговый метод).

- Выберите метод, при помощи которого будет отобрана та переменная, которая увеличивает расстояние Махаланобиса (Mahalanobis) между двумя группами. Эта дистанционная мера базируется на евклидовых расстояниях между нормализованными значениями выборок с учётом корреляции соответствующих переменных.
- Чтобы искусственно не раздувать объём выводимых результатов, в этот раз через кнопку Classify... (Классифицировать), активируйте опцию Summary table (Сводная таблица).

В рассматриваемом случае мы отказываемся от графического представления результатов. В анализ по очереди будут включены переменные: bzeit, gr, alter и kob; это те же самые переменные, которые использовались при применении метода логистической регрессии. По заключительной классификационной таблице можно сделать вывод о том, что в результате отбрасывания неподходящих переменных доля попаданий слегка выросла. Значение надёжности прогноза составило 70,2 %.

Для проведения дискриминантного анализа Вы можете использовать и пример с двумя диагностическими тестами для обнаружения карциномы мочевого пузыря, рассмотренный в главе 16.4. Здесь можно получить более чёткое разделение двух групп (здоров — болен). Точность прогнозирования здесь составляет 82,2 %.

## 18.2. Пример из области социологии

В своём исследовании "Культурный прорыв. Изменение ценностей в западном мире" (см. дополнительную литературу) Рональд Инглехарт (Ronald Inglehart) приводит тезис, что в более зрелых возрастных группах значимо большее количество человек высказались в пользу материальных ценностей (см. гл. 8.4.2). Среди младших поколений, согласно Инглехарту, растёт доля постматериалистов. Склонность опрошенных к постматериалистическим ценностям зависит от их образования и профессиональной квалификации. Чем выше образование и профессиональная квалификация, тем выше склонность к постматериалистическим ценностям. Значение имеет также и социально-экономический статус отца; согласно мнению Инглехарта, чем он выше, тем значительней доля постматериалистов. При помощи дискриминантного анализа мы проверим эту теорему смены ценностей, сформулированную американским политологом.

- Откройте в редакторе данных файл postmat.sav.

Переменные, которые вы сможете найти в этом файле, приводятся в нижеследующей таблице.

Имя переменной	Значение
ingMnd	Индекс Инглехарта
	Ценности:
	1 Постматериалисты
	2 Постматериалисты смешанного типа
	3 Материалисты смешанного типа
	4 Материалисты
	5 Не могу дать ответ
	6 Нет данных
statpaps	Социально-экономический статус отца (индекс)
	Значения:
	1 Низкий
	5 Высокий
	8 Формируется в данный момент (отсутствующее значение)
	9 Безработный, в заключении, умер, пенсионер и т.д. (отсутствующее значение)
schule	Уровень образования опрашиваемых
	Значения:
	1 Без образования
	2 Начальная школа
	3 Незаконченное среднее
	4 Среднее
alter	Возраст опрашиваемых

	Значения:
	1 18 до 29 лет
	2 30 до 44 лет
	3 45 до 59 лет
	4 60 до 74 лет
	5 75 до 88 лет
	6 89 и старше
	9 Не указан (отсутствующее значение)
ausbild	Профессиональное образование опрашиваемых Значения: 0 Образование отсутствует (отсутствующее значение) 1 Краткосрочное образование 2 Ученик 3 Мастер/техник 4 Высшее образование

Прежде чем приступить к дискриминантному анализу, преобразуем сначала переменную `ingl_ind` к дихотомическому типу. Значения признаков: 1 ("Постматериалисты") и 2 ("Постматериалисты смешанного типа") должны быть включены в новое значение признака 1 ("Постматериалистические типы") переменной `ingMnd`, а значения признаков: 3 ("Материалисты смешанного типа") и 4 ("Материалисты") в новое значение признака 2-"Материалистические типы".

- Для этого в редакторе синтаксиса введите следующие команды:

```
RECODE ingl_ind (1,2 = 1) (3,4 = 2)
INTO ingl_dic. VARIABLE LABELS
  ingl_dic = "Inglehart-Index, dichotom".
VALUE LABELS
  ingl_dic 1 "Postmat. Typen"
  2 "Materialist.Typen".
EXECUTE.
```

- Вы можете также загрузить в редактор синтаксиса и файл `ingledic.sps`, в котором находятся эти команды.
- Пометьте команды и запустите программу щелчком на кнопке Run Current (Выполнить текущие команды).

В редакторе данных появится новая переменная `ingl_dic`. Теперь проведите дискриминантный анализ.

- Выберите в меню опции Analyze (Анализ) Classify (Классифицировать) Discriminant... (Дискриминантный анализ)
- Переменную `ingl_dic` поместите в поле групповых переменных.
- Щёлкните на выключателе Define Range... (Определить область) и в качестве минимального значения введите 1, а в качестве максимального значения 2.
- Переменные `statpaps`, `schule`, `alter` и `ausbild` поместите в список Independents (Независимые переменные). Оставьте метод ввода переменных Enter independents together (Независимые переменные вводить одновременно), установленный по умолчанию.

Диалоговое окно Discriminant Analysis (Дискриминантный анализ) должно теперь выглядеть так, как показано на рисунке 18.4.

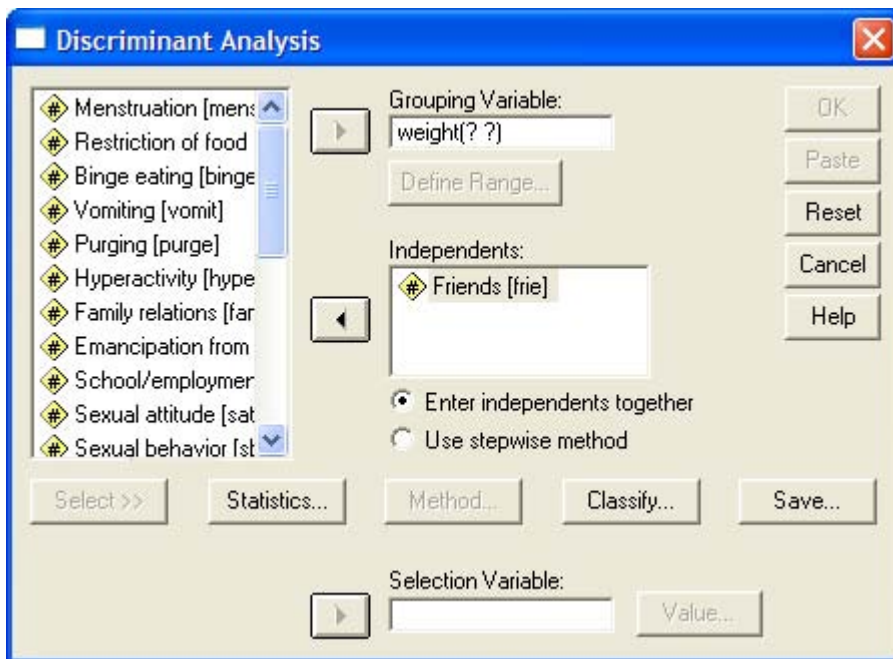


Рис. 18.4: Диалоговое окно *Discriminant Analysis* (Дискриминантный анализ).

- Щёлкните по выключателю *Statistics...* (Статистики)

Откроется диалоговое окно *Discriminant Analysis: Statistics* (Дискриминантный анализ: Статистики) (см. рис. 18.5).

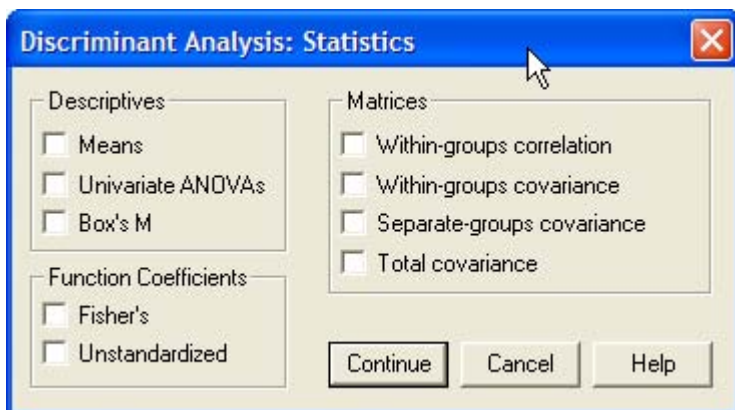


Рис. 18.5: Диалоговое окно *Discriminant Analysis: Statistics* (Дискриминантный анализ: Статистики)

- Активируйте опции: *Means* (Средние значения), *Univariate ANOVAs* (Одномерные тесты ANOVA), *Unstandardized Function Coefficients* (Не стандартизированные коэффициенты функции) и *Within-groups Correlation Matrices* (Корреляционная матрица внутри группы).
- Подтвердите нажатием *Continue* (Далее).
- Щёлкните на выключателе *Classify...* (Классифицировать). Откроется диалоговое окно *Discriminant Analysis: Classification* (Дискриминантный анализ: Классификация) (см. рис. 18.6).

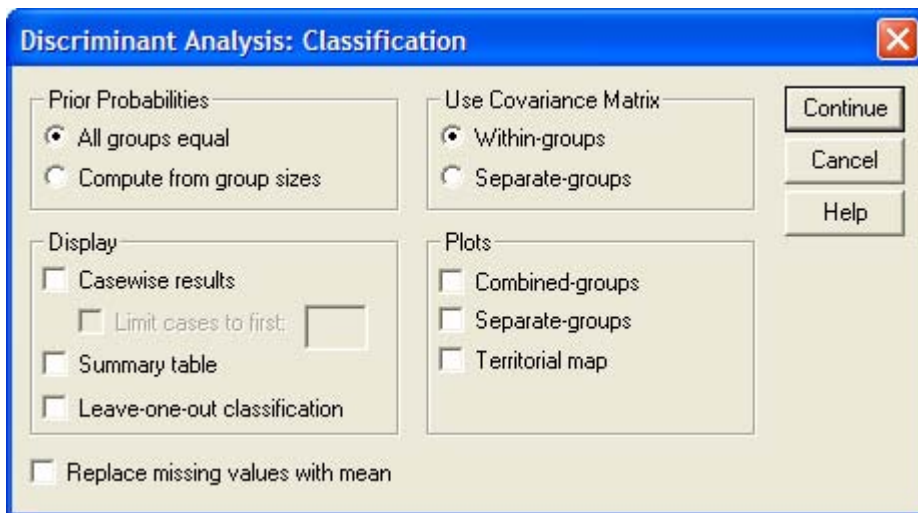


Рис. 18.6: Диалоговое окно *Discriminant Analysis: Classification* (Дискриминантный анализ: Классификация)

- Сделайте здесь запрос на Summary table (Сводную таблицу).
- Щёлкните на выключателе Save... (Сохранить). Откроется диалоговое окно *Discriminant Analysis: Save* (Дискриминантный анализ: Сохранить) (см. рис. 18.7).

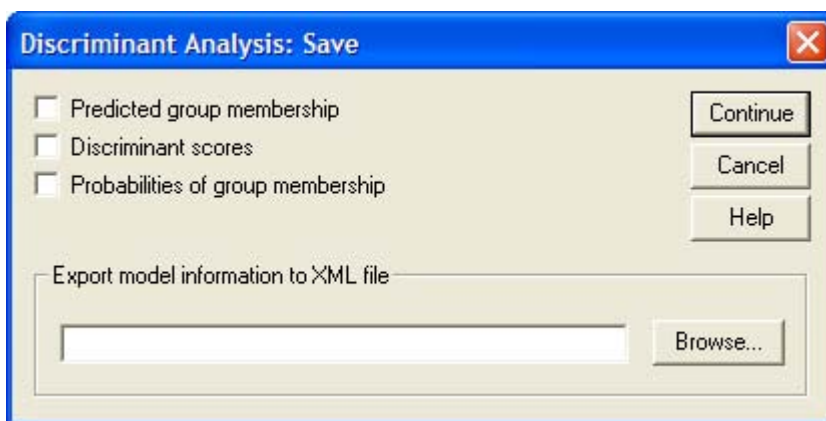


Рис. 18.7: Диалоговое окно *Discriminant Analysis: Save* (Дискриминантный анализ: Сохранить)

Видно, что в 10 версии появилась возможность сохранения информации о модели в так называемом, XML-файле (см. примечания к рис. 16.3).

- Активируйте вывод Predicted group membership (Прогнозируемой принадлежности к группе), Discriminant scores (Значений дискриминантной функции) и Probabilities of group membership (Вероятностей принадлежности к группе).
- Подтвердите нажатием Continue (Далее) и затем ОК.

В окне просмотра появится сначала обзор действительных и пропущенных значений:

#### Analysis Case Processing Summary (Анализ обработанных наблюдений)

Unweighted Cases (Не взвешенные случаи)		N	Percent (Процент)
Valid (Действительные)		2200	71,9
Excluded (Исключенные)	Missing or out-of-range group codes (Отсутствующие или находящиеся за пределами допустимой области кодировки принадлежности к группе)	19	,6

	At least one missing discriminating variable (По меньшей мере одна отсутствующая дискриминационная переменная)	816	26,7
	Both missing or out-of-range group codes and at least one missing discriminating variable (Обе кодировки принадлежности к группе отсутствуют или находятся за пределами допустимой области, или по меньшей мере одна отсутствующая дискриминационная переменная)	23	,8
	Total (Общее количество исключённых)	858	28,1
Total (Общее количество случаев)		3058	100,0

В общей сложности 858 наблюдений из 3058, находящихся в файле postmat.sav, были исключены из анализа из-за отсутствия значения переменной ingl\_dic или отсутствия значений одной из дискриминационных переменных. Таким образом анализ проводился для 2200 наблюдений. Далее приводятся средние значения, стандартные отклонения и количество наблюдений для всех переменных из обеих групп и для каждой группы в отдельности.

По средним значениям уже заметно, что для постматериалистических типов характерны: более высокий социально-экономический статус отца (2,8148 по сравнению с 2,3904), более высокое образование (2,9853 по сравнению с 2,5248) и принадлежность к младшей возрастной группе (2,1842 по сравнению с 2,8151).

### Group Statistics

<b>(Статистики для групп)</b>					
<b>INGL_DIC (Индекс Ингпехарта, дихото- мический)</b>		<b>Mean (сред- нее значе- ние)</b>	<b>Std. Deviation (Станда- ртное отклоне- ние)</b>	<b>Valid N (listwise) (Действительные значения (по списку))</b>	
				<b>Unwei- ghted (Не взвешен ные)</b>	<b>Weigh- ted (Взвешен ные)</b>
1,00 (Пост- материа- листический тип)	SES-Index des Vaters (социально- экономи- ческий статус отца)	2,8148	1,1718	1091	1091,000
	Schulabschluss (Образование)	2,9853	,8194	1091	1091,000
	ALTER, BEFRAGTE<R>, KATEGORISIERT (Возраст, опрошен- ного(ой), разбит на категории)	2,1842	1,0887	1091	1091,000
	Berufsaus-bildung (Профес- сиональное образо- вание)	2,1888	1,1562	1091	1091,000

2,00 (Материалистический тип)	SES-Index des Vaters (социально-экономический статус отца)	2,3904	1,0407	1109	1109,000
	Schulabschluss (Образование)	2,5248	,7627	1109	1109,000
	ALTER, BEFRAGTE<R>, KATEGORISIERT (Возраст , опрошенного(ой), разбит на категории)	2,8151	1,2111	1109	1109,000
	Berufsausbildung (Профессиональное образование)	1,8792	1,0249	1109	1109,000
Total (Сумма)	SES-Index des Vaters (социально-экономический статус отца)	2,6009	1,1275	2200	2200,000
	Schulabschluss (Образование)	2,7532	,8240	2200	2200,000
	ALTER, BEFRAGTE<R>, KATEGORISIERT (Возраст, опрошенного(ой), разбит на категории)	2,5023	1,1942	2200	2200,000
	Berufsausbildung (Профессиональное образование)	2,0327	1,1027	2200	2200,000

Затем проводится тест на значимость различия между переменными, относящимися к обеим группам, то есть выясняется присутствуют ли в них разделяющие (дискриминирующие) особенности, позволяющие судить об отношении к одной из двух групп (постматериалисты — материалисты).

#### Tests of Equality of Group Means (Тест равенства групповых средних значений)

	Wilks1 Lambda (Лямбда Уилкса)	F	df1	df2	Sig. (Значимость)
SES-Index des Vaters (социально-экономический статус отца)	,965	80,746	1	2198	,000
Schulabschluss (Образование)	,922	186,281	1	2198	,000
ALTER, BEFRAGTE<R>, KATEGORISIERT (Возраст, опрошенного(ых), разбит на категории)	,930	164,951	1	2198	,000
Berufsausbildung (Профессиональное образование)	,980	44,222	1	2198	,000

Как следует из колонки значимости, по всем переменным наблюдается значительное различие между группами ( $p < 0,001$ ).

Далее приводится корреляционная матрица между всеми переменными, причём коэффициенты были рассчитаны для обеих групп:

### Pooled Within-Groups Matrices (Объединённые матрицы внутри групп)

		SES-Index des Vaters (социально-экономический статус отца)	Schulabschluss (Образование)	ALTER, BEFRAG -TE<R>, KATEGORISIERT (Возраст, опрошенного(ой), разбит на категории)	Berufsausbildung (Профессиональное образование)
Correlation (Корреляция)	SES-Index des Vaters (социально-экономический статус отца)	1,000	,327	-,033	,137
	Schulabschluss (Образование)	,327	1,000	-,275	,377
	ALTER, BEFRAG -GTE<R>, KATEGORISIERT (Возраст, опрошенного(ых), разбит на категории)	-,033	-,275	1,000	,018
	Berufsausbildung (Профессиональное образование)	,137	,377	,018	1,000

Прежде всего, здесь очень заметна корреляция между переменными schule и statpas и между переменными ausbild и schule. Чем выше социально-экономический статус отца, тем выше школьное образование опрошиваемого; чем выше его школьное образование, тем выше и профессиональное образование.

Далее следует анализ коэффициентов дискриминантной функции. Корреляционный коэффициент между рассчитанными значениями дискриминантной функции и реальной принадлежностью к группе, равный 0,353, является неудовлетворительным:

### Eigenvalues (Собственные значения)

Function (Функция)	Eigenvalue (Собственное значение)	% of Variance (% дисперсии)	Cumulative % (Совокупный %)	Canonical Correlation (Каноническая корреляция)
1	,142a	100,0	100,0	,353

a. First 1 canonical discriminant functions were used in the analysis (Первые 1 канонические дискриминантные функции будут применяться в анализе).

### Wilks' Lambda (Лямбда Уилкса)

Test of Function(s) Wilks' Lambda (Тест функции (и)) (Лямбда Уилкса)	Chi-square (Хи-квадрат)	df	Sig. (Значимость)
1 ,875	292,431	4	,000

Тест, проведенный с помощью критерия "Лямбда Уилкса" (k), на предмет, значимо ли различаются между собой средние значения дискриминантной функции в обеих группах, показал очень значимый результат (значение  $p < 0,001$ ).

Затем приводятся стандартизированные коэффициенты дискриминантной функции и их корреляция с используемыми переменными:



## Standardized Canonical Discriminant Function Coefficients

<b>(Стандартизированные канонические коэффициенты дискриминантной функции)</b>	
	<b>Function (Функция)</b>
	1
SES-Index des Vaters (социально-экономический статус отца)	,321
Schulabschluss (Образование)	,434
ALTER, BEFRAGTE<R>, KATEGORISIERT (Возраст, опрошенного(ой), разбит на категории)	-,599
Berufsausbildung (Профессиональное образование)	,179

## Structure Matrix

<b>(Структурная матрица)</b>	
	<b>Function (Функция)</b>
	1
Schulabschluss (Образование)	,771
ALTER, BEFRAGTE<R>, KATEGORISIERT (Возраст, опрошенного(ой), разбит на категории)	-,726
SES-Index des Vaters (социально-экономический статус отца)	,508
Berufsausbildung (Профессиональное образование)	,376

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions (Объединённые корреляции внутри групп между дискриминантными переменными и стандартизированными каноническими дискриминантными функциями)

Variables ordered by absolute size of correlation within function (Переменные расположены соответственно величине их абсолютных корреляционных показателей).

После этого приводятся нестандартизированные коэффициенты дискриминантной функции и средние значения дискриминантной функции в обеих группах:

## Canonical Discriminant Function Coefficients

<b>(Канонические коэффициенты дискриминантной функции)</b>	
	<b>Function (Функция)</b>
	1
SES-Index des Vaters (социально-экономический статус отца)	,290
Schulabschluss (Образование)	,549
ALTER, BEFRAGTE<R>, KATEGORISIERT (Возраст, опрошенного(ой), разбит на категории)	-,520
Berufsausbildung (Профессиональное образование)	,164
(Constant) (Постоянно)	-1,297

Unstandardized coefficients (нестандартизированные коэффициенты)

## Functions at Group Centroids (Функции для групповых центроидов)

INGL DIC	Function (Функция)
	1
1 ,00 (Постматериалистический тип)	,380
2,00 (Материалистический тип)	-.374

Unstandardized canonical discriminant functions evaluated at group means (Нестандартизированные канонические дискриминантные функции, оценка которых происходит относительно средних значений групп).

В данном случае мы отказались от вывода очень длинной таблицы, в которой для каждого наблюдения построчно, приводится информация о значении дискриминантной функции и принадлежности к одной из двух групп.

В заключении приводится классификационная таблица с указанием точности попадания прогнозов:

### Classification Results<sup>a</sup>

(Классификационные результаты)					
		INGL_DIC (Индекс Инглехарта, дихото- мический)	Predicted Group Membership (Прогнозируемая принадлежность к группе)		Total (Сум- ма)
			1,00 (Постматериал тип)	2,00 (Материал тип)	
Original (Перво- начально)	Count (Ко- во)	1 ,00 (Постматериал тип)	710	381	1091
		2,00 (Материал тип)	410	699	1109
		Ungrouped cases (Не сгруппированные наблюдения)	7	12	19
	%	1 ,00 (Постматериал тип)	65,1	34,9	100,0
		2,00 (Материал тип)	37,0	63,0	100,0
		Ungrouped cases (Не сгруппированные наблюдения)	36,8	63,2	100,0

a. 64,0% of original grouped cases correctly classified (64 % наблюдений, первоначально разнесённых по группам, были классифицированы корректно).

Правая колонка таблицы ("Total" (Сумма)) указывает на общее количество наблюдений, которые фактически относятся к соответствующим группам. К группе постматериалистических типов относится 1091 наблюдение, а к группе материалистических типов 1109. Обе колонки, объединённые общим наименованием ("Predicted Group Membership" (Прогнозируемая принадлежность к группе)), указывают на фактическое количество наблюдений, относящихся к каждой из групп. Первая колонка указывает на количество наблюдений, которые были отнесены к первой группе. Из 1091 постматериалистических наблюдений корректно определены были 710, это соответствует 65,1 % всех наблюдений. 381 наблюдение было по ошибке отнесено ко 2 группе, что соответствует 34,9 % всех наблюдений. Из 1109 материалистических наблюдений по ошибке к группе 1 были отнесены 410, что соответствует 37,0 %. 699 наблюдений были корректно отнесены к группе 2, что составило 63 %. Строка "Ungrouped cases" (Несгруппированные наблюдения) содержит наблюдения, которые не соответствуют ни одной из

групп. Хотя эти наблюдения и не учитываются при расчёте дискриминантной функции, значение функции для них всё равно вычисляется. Из 19 наблюдений, для которых отсутствуют данные о принадлежности к какой-либо группе, 7 были отнесены к постматериалистическим типам, а 12 к материалистическим. В строке под таблицей приводится итоговый результат. 64 % наблюдений были классифицированы корректно. Так как даже при чисто случайном отнесении некоторого наблюдения к одной из двух имеющихся групп, корректность классификации данного наблюдения составила бы 50 %, то 64 %-ную точность прогноза следует рассматривать как довольно умеренный результат. Такой неудовлетворительный результат можно попытаться объяснить тем, что в обе группы входили смешанные типы, которые тяжелее классифицировать, нежели чистые типы. Проверим это предположение путём повторного проведения расчёта, но уже с учётом только чистых типов.

- Выберите в меню Data (Данные) Select Cases... (Выбрать наблюдения)
- Щёлкните на опции If condition is satisfied (Если выполняется условие) и затем на выключателе If... (Если).
- В редакторе условий введите следующее условие:

ing1\_ind = 1 OR ing1\_ind = 4

- Подтвердите нажатием Continue (Далее) и затем ОК.
- В диалоговом окне Discriminant Analysis (Дискриминантный анализ) переменную ing1\_ind (не ing1\_dic!) поместите в поле для групповых переменных. В качестве границ области изменения задать значения 1 и 4.
- В список независимых переменных поместите переменные statpaps, schule, alter и ausbild.
- Дополнительные установки под выключателями Statistics... (Статистики), Classify... (Классифицировать) и Save... (Сохранить) произведите так, как было описано ранее.

Вы получите следующую классификационную таблицу:

### Classification Results

<b>(Результаты классификации)</b>					
		<b>INGLEHART-INDEX (Индекс Инглехарта, дихотомический)</b>	<b>Predicted Group Membership (Прогнозируемая принадлежность к группе)</b>		<b>Total (Сумма)</b>
			<b>POSTMATE-RIALISTEN (Постматериал)</b>	<b>MATERI-ALISTEN (Материал)</b>	
Original (Первоначально)	Count (Кол-во)	POSTMATE-RIALISTEN (Постматериал)	409	109	518
		MATERI-ALISTEN (Материалисты)	133	297	430
	%	POSTMATE-RIALISTEN (Постматериал)	79,0	21,0	100,0
		MATERI-ALISTEN (Материалисты)	30,9	69,1	100,0

a. 74,5% of original grouped cases correctly classified (74,5 % наблюдений, первоначально разнесённых по группам, были классифицированы корректно).

К группе постматериалистов относится 518 наблюдений. 409 наблюдений (79 %) были спрогнозированы корректно, а 109 (21,0 %) по ошибке отнесены к группе 4 ("чистые материалисты"). В группе чистых материалистов насчитывается 403 наблюдения. 297 наблюдений (69,1 %) были определены корректно, а 133 (30,9 %) по ошибке были отнесены к

группе 1 ("чистые постматериалисты"). Конечным результатом является корректная идентификация наблюдений, равная 74,5 %. Этот показатель значительно выше предыдущего и может быть расценен как приемлемый.

### 18.3. Пример из области биологии

Дискриминантный анализ очень часто применяется для обработки данных из области биологии. В следующем типичном примере для некоторого количества индивидуумов принадлежность к группе уже известна, на основании чего и строится дискриминантная функция. Далее она используется для того, чтобы оценить принадлежность к определенной группе тех индивидуумов, для которых она ещё не известна.

В файле `vogel.sav` хранятся данные о половой принадлежности, длине крыла, длине клюва, размере головы, длине лап и весе 245 птиц определённого вида. Причём пол смогли определить только для 51 особи. Кодировка пола соответствует 1 = мужской и 2 = женский; отсутствие данных кодируется 9.

Если для перечисленных параметров Вы рассчитаете средние значения для самцов и самок, то для самок получите более высокие показатели. Исходя из этого, при помощи дискриминантного анализа можно попытаться определить пол тех особей, для которых этого нельзя было сделать ранее.

- Откройте файл `vogel.sav`.
- В диалоговом окне Discriminant Analysis (Дискриминантный анализ) переменной `geschl` (Пол) присвойте статус групповой переменной с пределами от 1 до 2, а переменным `fluegel` (Длина крыла), `schnl` (Длина клюва), `korfl` (Размер головы), `fuss` (Длина лап) и `gew` (Вес) — статус независимых переменных. Выберите пошаговый метод.
- В диалоговом окне Discriminant Analysis: Classify (Дискриминантный анализ: Классифицировать) активируйте Casewise results (Результаты для отдельных наблюдений) с ограничением в 40 наблюдений и Summary table (Сводная таблица).
- Через выключатель Save... (Сохранить) при помощи активирования опций Predicted group membership (Прогнозируемая принадлежности к группе) и Probabilities of group membership (Вероятности принадлежности к группе) затребуйте генерирование соответствующих переменных.

Из всех результатов, приводимых в окне просмотра, в книге рассматриваются только статистики для каждого наблюдения. По классификационной таблице видно, что для 51 наблюдения с заранее известным полом 44 раза, т.е. в 86,3 % наблюдений, пол был спрогнозирован верно (см. следующую таблицу).

Если мы рассмотрим наблюдение 8, то здесь пол известен — женский и в результате прогноза получается женский пол, а вот для наблюдения 30 пол известен как мужской, но прогнозируется как женский. Наблюдения с нераспознанным полом приводятся в таблице как "ungrouped" (не группированные).

Для наблюдения 1, для которого пол оказался неизвестным, он прогнозируется как женский. Значение вероятности прогнозирования, 0,990, указывается в колонке "P(G=g | D=d)" под заголовком "Highest Group" (Старшая группа). Менее достоверным является прогноз пола для наблюдения 10, здесь вероятность прогнозирования составляет только 0,721.

#### **\*\* Misclassified case (\*\* - Неверно классифицированный случай)**

Для того, чтобы хотя бы частично сократить количество ошибочных значений для переменной пола, при анализе вы можете применять прогнозируемую групповую принадлежность только в тех случаях, для которых вероятность прогнозирования принимает некоторое минимально допустимое значение, к примеру, 0,9.

```

IF (dis_1 = 1 and disl_1 >= 0,9)
  geschl=1
. IF (dis_1 = 2 and dis2_1 >= 0,9)
  geschl=2.
EXECUTE.

```

Таким образом, в используемом примере можно присвоить половой показатель ещё 90-а птицам. Если вы снизите минимально допустимое значение вероятности прогнозирования, то это число станет ещё больше.

К файлу были добавлены три новые переменные:

dis\_1: Прогнозируемая группа

disl\_1: Вероятность принадлежности к группе 1

dis2\_1: Вероятность принадлежности к группе 2.

## 18.4. Пример из области биологии (три группы)

В предыдущих примерах дискриминантный анализ всегда проводился при наличии лишь двух групп. В этой главе рассматривается пример, в котором групповая переменная имеет больше двух категорий, а именно три.

В файле *kaefer.sav* содержатся данные о длине и ширине грудной клетки трёх видов жуков (обозначенных как А, В и С). Если вы проведёте однофакторный дисперсионный анализ с последующими дополнительными тестами (Post-hoc-Tests), то увидите, что три разновидности жуков очень значительно различаются между собой как по длине, так и по ширине, поэтому вполне можно предположить, что этих жуков можно классифицировать между упомянутыми видами на основании их длины и ширины посредством дискриминантного анализа.

- Откройте файл *kaefer.sav*.

Вы увидите, что 17 жуков из 30 не отнесены ни к одной из групп; поэтому классификация жуков по группам должна быть произведена при помощи дискриминантного анализа.

- В диалоговом окне Discriminant Analysis (Дискриминантный анализ) переменной *kaefer* (Жук) присвойте статус групповой переменной с пределами от 1 до 3, а переменным *laenge* (Длина) и *breite* (Ширина) статус независимых переменных. Оставьте активной установку по умолчанию *Enter independents together* (Независимые переменные вводить одновременно).
- В диалоговом окне Discriminant Analysis: Statistics (Дискриминантный анализ: Статистики) в разделе Descriptives (Дискриптивные статистики) активируйте опции: Means (Средние значения), Univariate ANOVAs (Одномерные тесты ANOVA) и в разделе Function Coefficients (Коэффициенты функции) опцию Unstandardized (Не стандартизированные).
- В диалоговом окне Discriminant Analysis: Classify (Дискриминантный анализ: Классифицировать) сделайте запрос на Case-wise results (Результаты для отдельных наблюдений) и Summary table (Сводную таблицу) и в разделе Plots (Графики) активируйте опцию Territorial map (Территориальная карта). Эта опция служит для построения классификационной диаграммы, так называемой территориальной карты (Territorial map). Построение этой диаграммы типично для случая с более чем двумя группами.
- В заключение, в диалоговом окне Discriminant Analysis: Save (Дискриминантный анализ: Сохранить), активируйте все опции, находящиеся там, с целью создания соответствующих переменных в исходном файле.

Из всей гаммы приводимых результатов расчёта мы рассмотрим только самые важные. Из групповых статистик можно узнать, что в семейство А входят самые большие, а в семейство В самые маленькие жуки.

## Group Statistics

<b>(Статистики для групп)</b>					
<b>КАЕФЕР (Жук)</b>		<b>Mean (Среднее значение)</b>	<b>Std. Deviation (Стандартное отклонение)</b>	<b>Valid N (listwise) (Действительные значения по списку))</b>	
				<b>Unweighted (Не взвешенное)</b>	<b>Weighted (Взвешенное)</b>
1 (Семейство А)	LAENGE (Длина)	1,6226	5.968E-02	42	42,000
	BREITE (Ширина)	1,2607	4J54E-02	42	42,000
2 Семейство В)	LAENGE (Длина)	1,3089	7.634E-02	45	45,000
	BREITE (Ширина)	1,0122	4.415E-02	45	45,000
3 Семейство С)	LAENGE (Длина)	1,4788	6.029E-02	26	26,000
	BREITE (Ширина)	1,1192	5.114E-02	26	26,000
Total	LAENGE (Длина)	1,4646	,1535	113	113,000
	BREITE (Ширина)	1,1292	,1191	113	113,000

Статистика Лямбда Уилкса (>i) свидетельствует о том, что жуки очень значимо делятся на группы как по длине, так и по ширине.

### Tests of Equality of Group Means (Тест на равенство средних значений групп)

	<b>Wilks' Lambda (Лямбда Уилкса)</b>	<b>F</b>	<b>df1</b>	<b>df2</b>	<b>Sig. (Значимость)</b>
LAENGE (Длина)	,187	239,154	2	110	,000
BREITE (Ширина)	,153	303,326	2	110	,000

Если насчитывается более двух классификационных групп, то можно образовать больше одной дискриминантной функции; при трёх группах, как в приведенном примере, их будет две. Следующая таблица свидетельствует о том, что обе дискриминантные функции дают значимые результаты для разделения между группами и, следовательно, могут быть использованы соответствующим образом. Однако, первая функция дает вероятность прогноза 98,7 %, а вторая только 1,3 %.

### Eigenvalues (Собственные значения)

<b>Function (Функция)</b>	<b>Eigenvalue (Собственные значение)</b>	<b>% of Variance (% дисперсии)</b>	<b>Cumulative % (Совокупный %)</b>	<b>Canonical Correlation (Каноническая корреляция)</b>
1	6,040a	98,7	98,7	,296
2	,078a	1,3	100,0	,269

a. First 2 canonical discriminant functions were used in the analysis (В этом анализе используются первые 2 канонические дискриминантные функции).

### Wilks' Lambda (Лямбда Уилкса)

Test of Function(s) (Тест функции (й))	Wilks1 Lambda (Лямбда Уилкса)	Chi-square (Хи-квадрат)	df	Sig. (Значимость)
1 through 2 (1 до 2)	,132	221,900	4	,000
2	,928	8,202	1	,004

Затребованные нестандартизированные коэффициенты функций приводятся в следующей таблице.

### Canonical Discriminant Function Coefficients

(Канонические коэффициенты дискриминантных функций)		
	Function (Функция)	
	1	2
LAENGE (Длина)	5,831	18,769
BREITE (Ширина)	14,891	-23,659
(Constant) (Константа)	-25,355	-,773

### Unstandardized coefficients (Нестандартизированные коэффициенты)

Мы здесь опускаем вывод статистик для каждого отдельного случая. В результате расчетов Вы получаете соответствующие номера групп и вероятность прогнозирования под заголовком  $P(G = g|D = d)$ . Прогнозирование осуществлено и для 17 неклассифицированных случаев.

На территориальной карте показано разделение на области, которые означают принадлежность к группе. При этом в пределах границ соответствующей области вероятность отнесения к данной группе выше, чем для других групп. На границах областей вероятности для граничащих групп одинаковы.

Значения обеих дискриминантных функций, на основе которых построена эта территориальная карта, Вы можете увидеть в редакторе данных под именами двух вновь созданных переменных: dis1\_1 и dis2\_1.

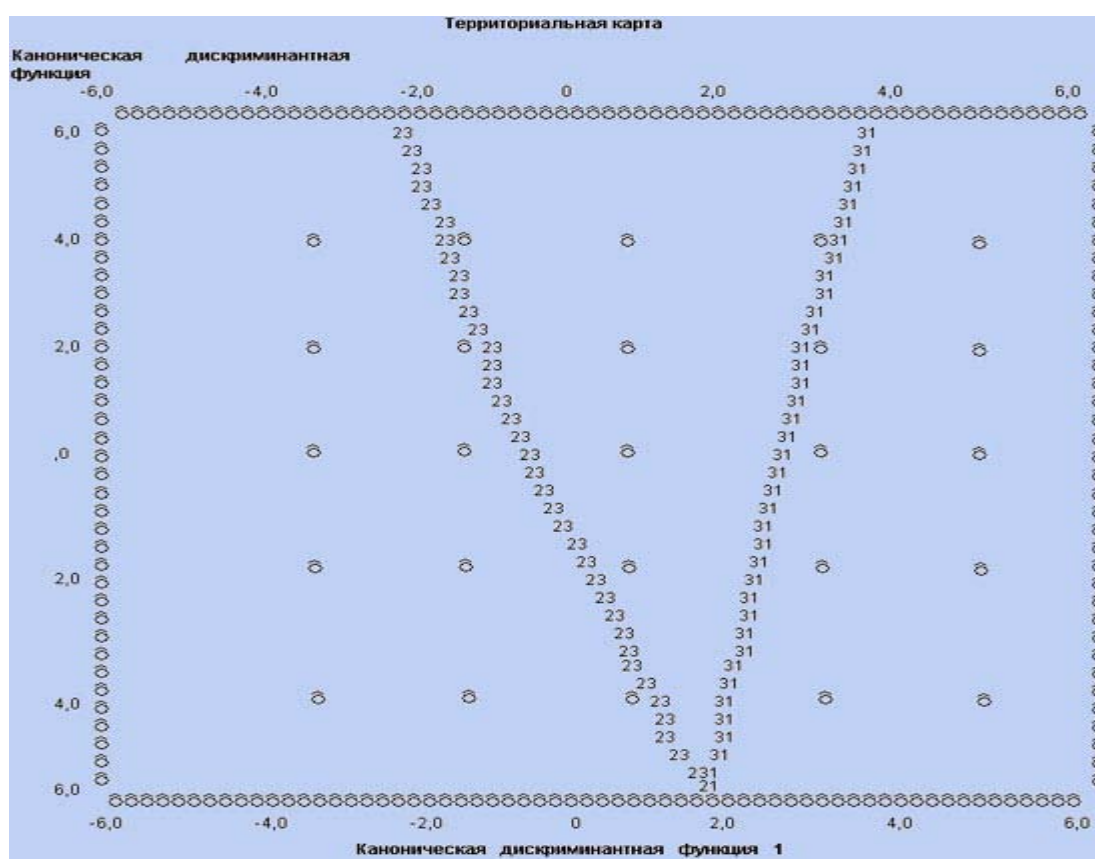
В заключение приводится обзор результатов классификации. По ним Вы можете заметить, что прогноз для групп А и В практически полностью был сделан верно и корректно классифицированы, в общей сложности, 91,2 % всех случаев.

### Classification Results <sup>a</sup>

(Результаты Классификации)						
		FUND (Семейство)	Predicted Group Membership			Total (Сумма)
			1 (Семейство А)	2 (Семейство В)	3 (Семейство С)	
Original (Перво- начально)	Count (Кол-во)	1 (Семейство А)	41	0	1	42
		2 (Семейство В)	0	43	2	45
		3 (Семейство С)	4	3	19	26

		Ungrouped cases (Не группированные случаи)	7	6	4	17
	%	1 (Семейство А)	97,6	,0	2,4	100,0
		2 (Семейство В)	,0	95,6	4,4	100,0
		3 (Семейство С)	15,4	11,5	73,1	100,0
		Ungrouped cases (Не группированные случаи)	41,2	35,3	23,5	100,0

а. 91,2% of original grouped cases correctly classified (91,2 % первоначально сгруппированных случаев были классифицированы корректно).



Символы, используемые в территориальной карте		
Символ	Группа	Метка
1	1	Семейство А
2	2	Семейство В
3	3	Семейство С
Маркировка		Центроиды групп

Наряду с уже упоминавшимися значениями обеих дискриминантных функции в редакторе данных были созданы: переменная *dis\_1*, содержащая значение прогнозируемой группы и переменные *dis1\_2*, *dis2\_2* и *dis3\_2*, которые содержат прогнозируемые вероятности отнесения к одной из трёх групп. Группа, которой соответствует наибольшая вероятность прогнозирования и есть прогнозируемая группа.