

Дискриминантный анализ

- Назначение дискриминантного анализа
- Виды ДА
- Сходство и различия между ДА и регрессионным и дисперсионным анализом
- Алгоритм ДА

Дискриминантный анализ

Основная идея:

- Мы измерили целый набор переменных, **ИЗНАЧАЛЬНО ЕСТЬ ГРУППЫ.**
- Мы хотим понять, **чем отличаются** между собой эти группы (на основе данных переменных).
- (Когда потом мы измерим эти переменные у **новой особи**, мы сможем с известной вероятностью **отнести её к той или иной группе**).

Дискриминантный анализ

Изучаем лемуров на Мадагаскаре.

Есть 3 вида лемуров, мы поймали зверьков разных видов, взвесили, померили длину черепа и резцов.

Вопрос: можем ли мы отличить виды по этим переменным?



Назначение

- С помощью ДА на основании некоторых признаков объект может быть причислен к одной из двух или нескольких групп.
- В результате анализа строится **дискриминантная функция**

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

где x_1, \dots и x_n — значения переменных, соответствующих рассматриваемым случаям,

b_1-b_n и b_o — коэффициенты, которые нужно оценить.

Коэффициенты подбираются так, чтобы по значениям дискриминантной функции можно было с максимальной четкостью провести разделение по группам

Переменные в ДА

Дискриминантный анализ используется для анализа данных в том случае, когда зависимая переменная категориальная, а предикторы (независимые переменные) - интервальные.

Пример

зависимая переменная - **выбор торговой марки персонального** компьютера (торговые марки А, В или С)

независимые переменные - **рейтинги свойств персональных компьютеров**, измеренные по семибалльной шкале Лайкерта

Виды ДА

- **Дискриминантный анализ для двух групп (two-group discriminant analysis)**
 - зависимая переменная имеет две категории.
- **Множественный дискриминантный анализ (multiple discriminant analysis)**
 - у зависимой переменной имеется три или больше категорий.

Примеры ДА в маркетинге

- Чем, с точки зрения демографических характеристик, отличаются приверженцы данного магазина от тех, у кого эта приверженность отсутствует?
- Отличаются ли в потреблении замороженных продуктов покупатели, которые пьют безалкогольные напитки мало, умеренно и много?
- Какие психографические характеристики помогают провести различия между восприимчивыми и не восприимчивыми к цене покупателями компьютерной техники?
- Различаются ли между собой различные сегменты рынка по своим предпочтениям к средствам массовой информации?

Сходства и отличия между дисперсионным, регрессионным и дискриминантным анализом

	<i>Дисперсионный анализ</i>	<i>Регрессионный анализ</i>	<i>Дискриминантный анализ</i>
<i>Сходства</i>			
Число зависимых переменных	Одна или несколько	Одна	Одна
Число независимых переменных	Несколько	Несколько	Несколько
<i>Отличия</i>			
Природа зависимой переменной	Метрическая	Метрическая	Категориальная
Природа независимой переменной	Категориальная	Метрическая	Метрическая

Статистики, используемые в ДА

Каноническая корреляция (*canonical correlation*). Измеряет степень связи между дискриминантными показателями и группами. Это мера связи между единственной дискриминирующей функцией и набором фиктивных переменных, которые определяют принадлежность к данной группе.

Центроид (средняя точка) (*centroid*). это средние значения для дискриминантных показателей конкретной группы. Центроидов столько, сколько групп, т.е. один центроид для каждой группы. Средние группы для всех функций — это *групповые центроиды*.

Классификационная матрица (*classification matrix*). Иногда ее называют *смешанной матрицей*, или *матрицей предсказания*. Содержит ряд правильно классифицированных и ошибочно классифицированных случаев

Статистики, используемые в ДА

Коэффициенты дискриминантной функции (*discriminant function coefficients*). (ненормированные) — коэффициенты переменных, когда они измерены в первоначальных единицах.

Дискриминантные показатели (*discriminant scores*). Сумма произведений ненормированных коэффициентов дискриминантной функции на значения переменных, добавленная к постоянному члену.

Собственное (характеристическое) значение (*eigenvalue*). Для каждой дискриминантной функции собственное значение — это отношение межгрупповой суммы квадратов к внутригрупповой сумме квадратов. Большие собственные значения указывают на функции более высокого порядка.

Статистики, используемые в ДА

f-статистика и ее значимость (*f values and their significance*).

Значения статистики вычисляет однофакторный дисперсионный анализ, разбивая на группы по независимым переменным. Каждый предиктор, в свою очередь, служит в ANOVA метрической зависимой переменной.

Средние группы и групповые стандартные отклонения (*group means and group standard deviations*). Эти показатели вычисляют для каждого предиктора каждой группы.

Объединенная межгрупповая корреляционная матрица (*pooled within-group correlation matrix*). Объединенную межгрупповую корреляционную матрицу вычисляют усреднением отдельных ковариационных матриц для всех групп.

Статистики, используемые в ДА

- **Нормированные коэффициенты дискриминантных функций** (*standardized discriminant function coefficients*). Коэффициенты дискриминантных функций используют как множители для нормированных переменных, т.е. переменных с нулевым средним и дисперсией, равной 1.
- **Структурные коэффициенты корреляции** (*structure correlations*). Также известны как *дискриминантные нагрузки*, представляют собой линейные коэффициенты корреляции между предикторами и дискриминантной функцией.
- **Общая корреляционная матрица** (*total correlation matrix*). Если при вычислении корреляций наблюдения обрабатывают так, как будто они взяты из одной выборки, то в результате получают общую корреляционную матрицу.
- .

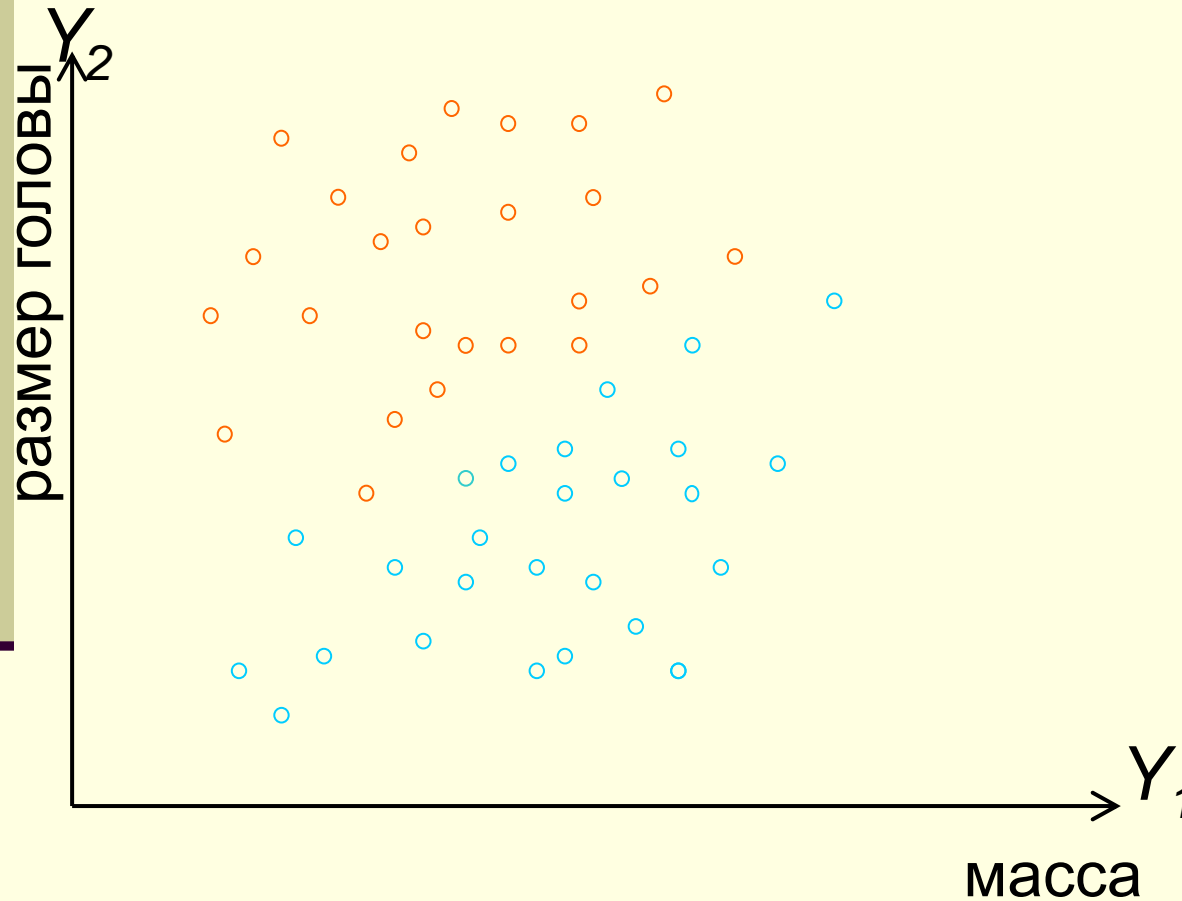
Статистики, используемые в ДА

- **Коэффициент лямбда Уилкса (Wilks's λ).** Иногда называемый f-статистикой, коэффициент Уилкса для каждого предиктора — это отношение внутригрупповой суммы квадратов к общей сумме квадратов.
- Его значение варьирует от 0 до 1.
- Большое значение λ (около 1) указывает на то, что средние групп не должны различаться.
- Малые значения (около 0) указывают на то, что средние групп различаются

Алгоритм ДА

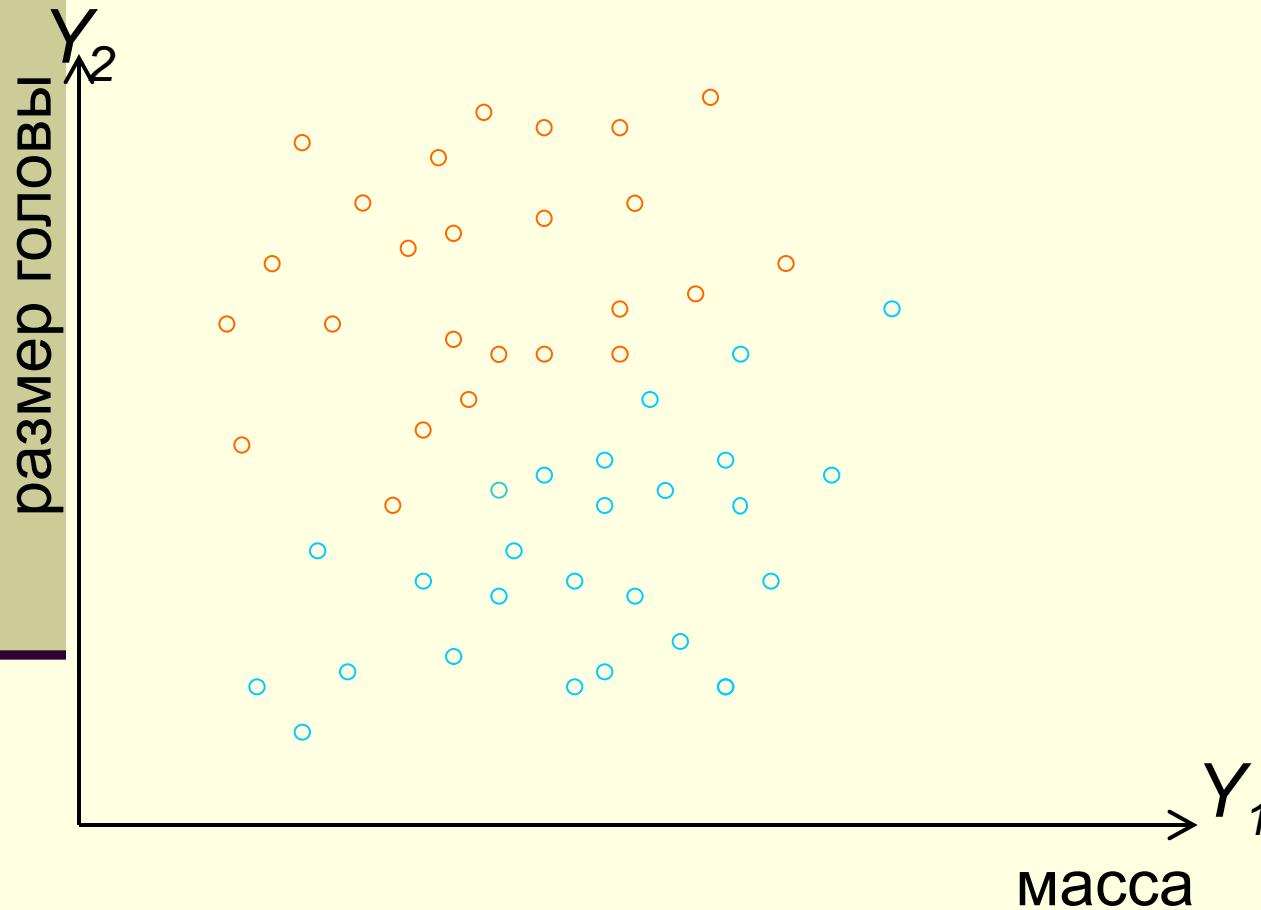
- формулирование проблемы
- разделение выборки на две части
- вычисление коэффициентов дискриминантной функции
- определение значимости различий между группами
- интерпретация и проверка достоверности

Лемуры



Средние значения для каждой переменной у разных видов отличаются, но их распределения перекрываются и для массы, и для головы, и для зубов!

Лемуры



Переменная Z
(**дискриминантная
функция**)

строится таким
образом, чтобы как
можно больше
зверьков одного из
видов получили
высокие значения Z ,
и как можно больше
зверьков другого
вида – низкие
значения Z .

Создание дискриминантной функции

Из выбранных переменных рассчитывают новые переменные Z (**дискриминантные функции**) – **линейные комбинации** исходных переменных, первая из которых наилучшим образом разделит группы (напр., виды).

Если группы **две**: получается одно уравнение.

Когда групп и исходных переменных много, получают **несколько дискриминантных функций** (всего $k-1$ или $p-1$ функций (k – число групп, p – число переменных, выбирают меньшее из этих чисел), «перпендикулярных» друг другу.

$$z_{ik} = b_1 y_{i1} + b_2 y_{i2} + \dots + b_j y_{ij} + \dots + b_p y_{ip}$$

На основе функции тестируют гипотезу о различии групп

Интерпретация дискриминантных функций

Каждую дискриминантную функцию характеризует **eigenvalue = Root** (собственное значение), и мы можем проверить, сколько функций при анализе действительно помогает различить группы, и какую часть изменчивости они объясняют (и исключить недостоверные).

standardized *b* coefficient = элементы **eigenvector**, b_j – позволяют оценить вклад каждой из переменных в данную дискриминантную функцию.

Структура факторов (**factor structure coefficients = loadings**) – позволяет понять, насколько какие переменные коррелируют с дискриминантными функциями.

Дискриминантный анализ

Теперь, когда мы построили такую функцию, мы сможем поймать зверька неизвестного вида, измерить у него Y_1 и Y_2 , рассчитать значение Z на основе уже посчитанных коэффициентов, и с некоторой точностью причислить его к тому или другому виду.



© 2006 Encyclopædia Britannica, Inc.

Discriminant function analysis

Data: лемуры.sta (9v by 58c)

	1	2	3	4	5
	омер лемур	вид	масса	голова	зуб верхний
1	#430	кошачий	1848	59,4	4,5
2	#74	чёрный	4249	89	5,5
3	#291	сифака	3444	86	5,4
4	#461	кошачий	2442	62,25	4,8
5	#210	чёрный	3787	83,4	5,5
6	#1044	кошачий	1968	58,05	5
7	#394	сифака	3822	85,8	5,5
8	#238	чёрный	4746	89	5,7
9	#130	чёрный	4956	91	5,4
10	#370	сифака	2849	87,6	5,2
11	#268	сифака	4564	88,6	6,2
12	#225	чёрный	3857	85,2	6
		чёрный	4347	88,2	5,5
		сифака	3045	84	5,5
		чёрный	5509	90,8	6
		чёрный	3976	88,8	5,5
		сифака	3185	85,6	5,8
		чёрный	3836	88	5,4
		чёрный	3437	87,2	5,5
		кошачий	1956	58,5	4,8
		чёрный	4298	92	5,5
		сифака	3535	84	5,3

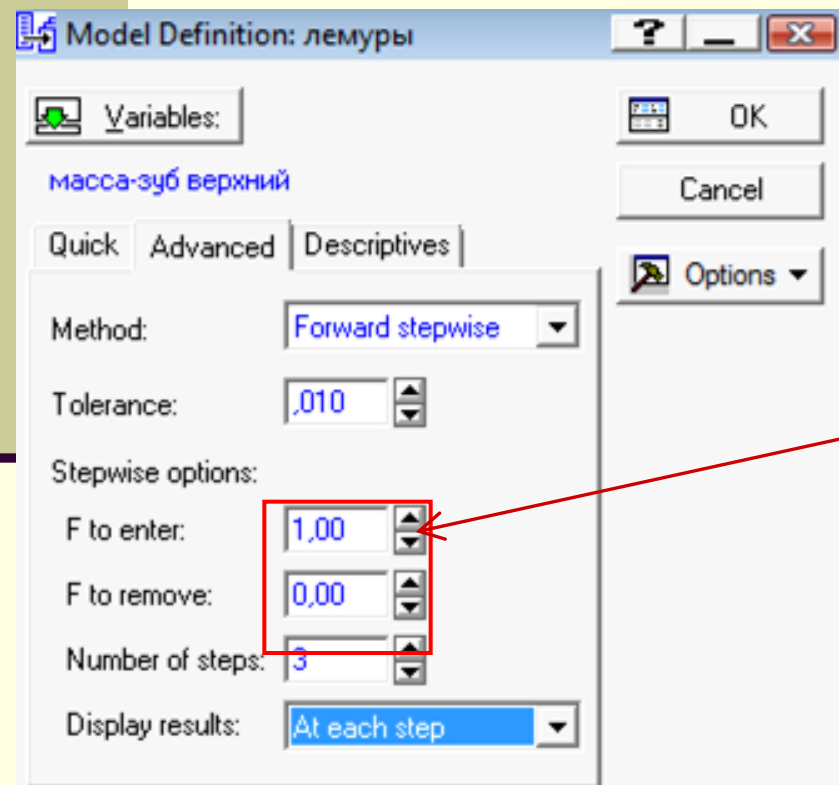
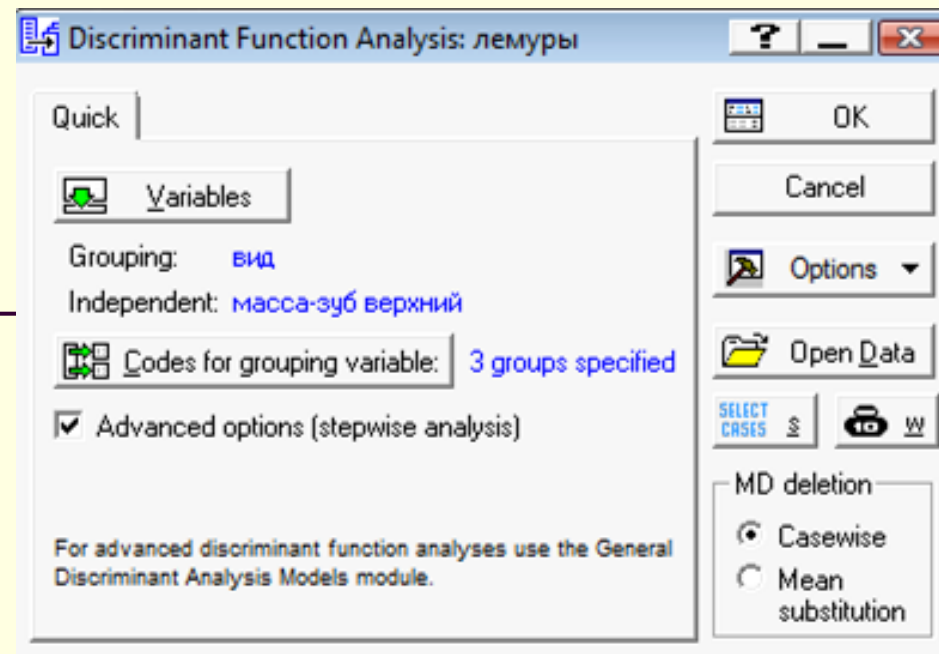
Resume... Ctrl+R

Add to Report

- Basic Statistics/Tables
- Multiple Regression
- ANOVA
- Nonparametrics
- Distribution Fitting
- Advanced Linear/Nonlinear Models
- Multivariate Exploratory Techniques**
 - Cluster Analysis
 - Factor Analysis
 - Principal Components & Classification Analysis
 - Canonical Analysis
 - Reliability/Item Analysis
 - Classification Trees
 - Correspondence Analysis
 - Multidimensional Scaling
 - Discriminant Analysis**
 - General Discriminant Analysis Models
- Industrial Statistics & Six Sigma
- Power Analysis
- Data-Mining
- Statistics of Block Data
- STATISTICA Visual Basic
- Probability Calculator

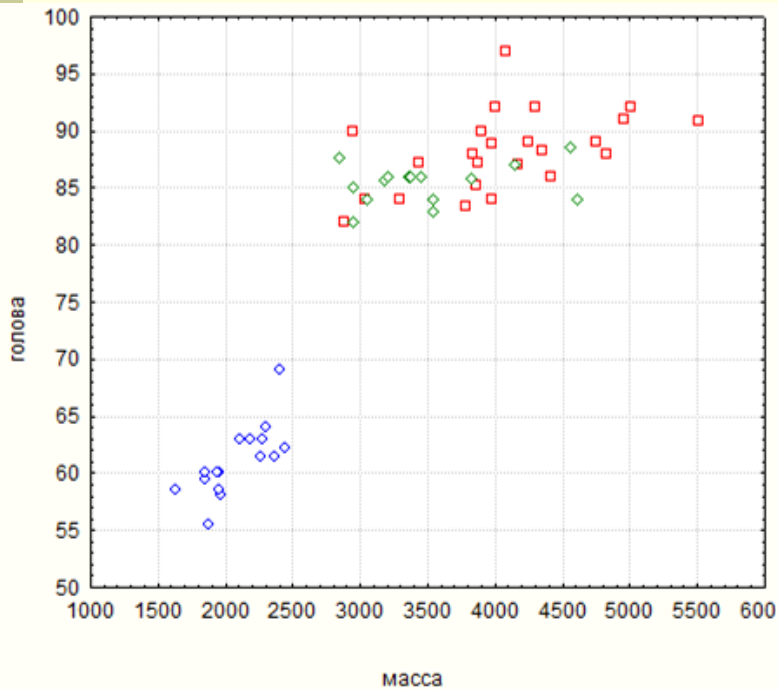
51	7	#394	сифака
52	8	#238	чёрный

Создание модели

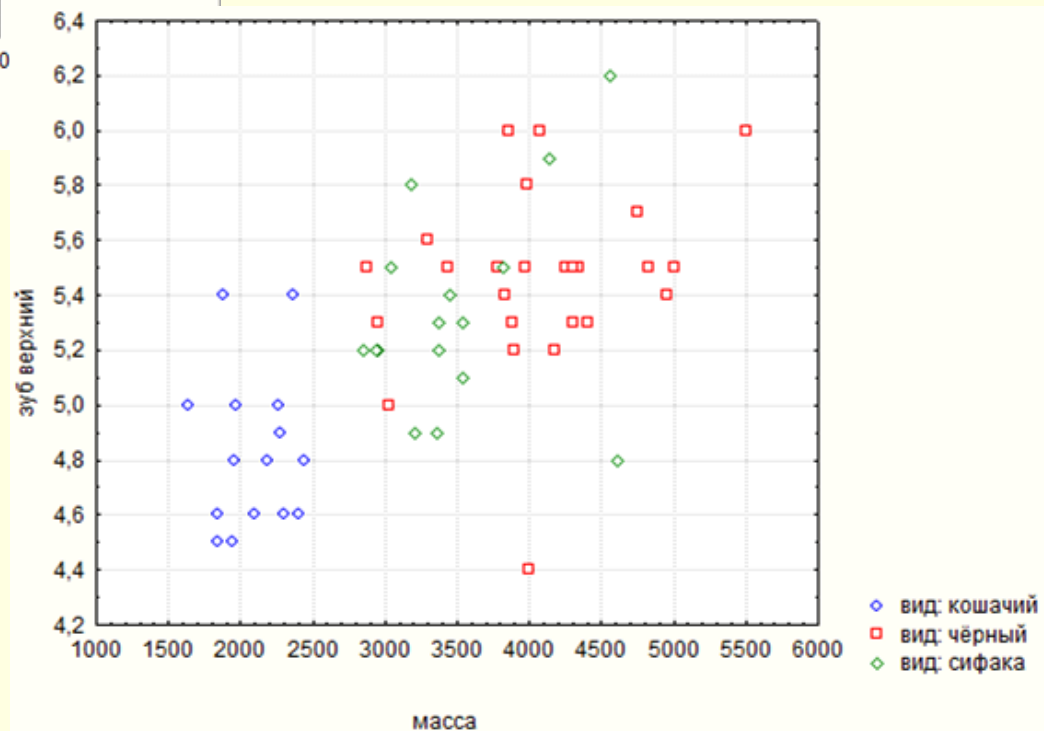


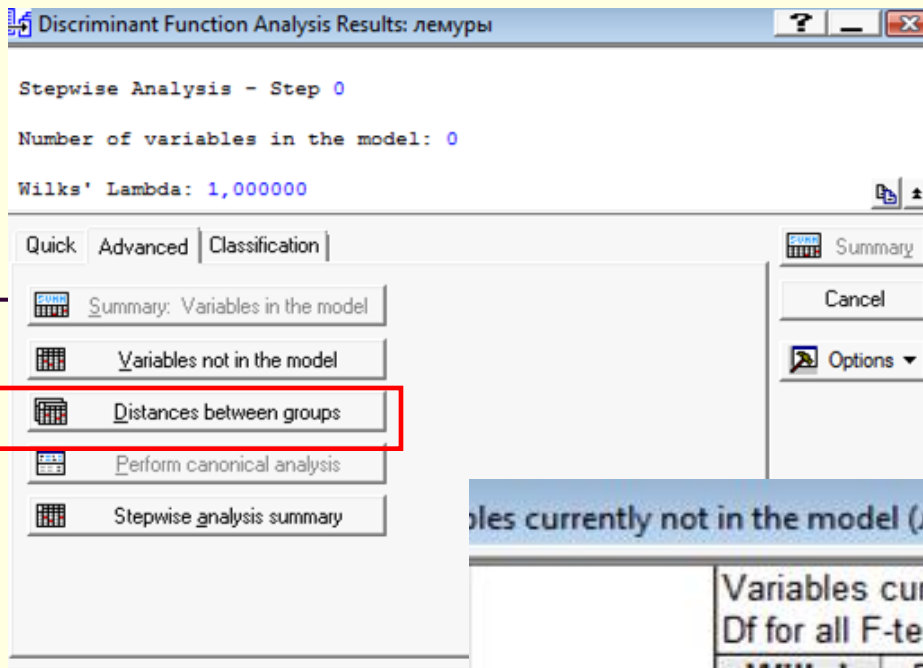
Критерии, по которым включают переменные для построения дискриминантной функции, задают минимальными.

Толерантность – $1 - R^2$, где R^2 показывает корреляцию данной переменной с остальными, т.е., позволяет исключить избыточные переменные.



Прежде чем приступить к анализу, смотрят, есть ли разделение на группы по переменным (аналог ANOVA).

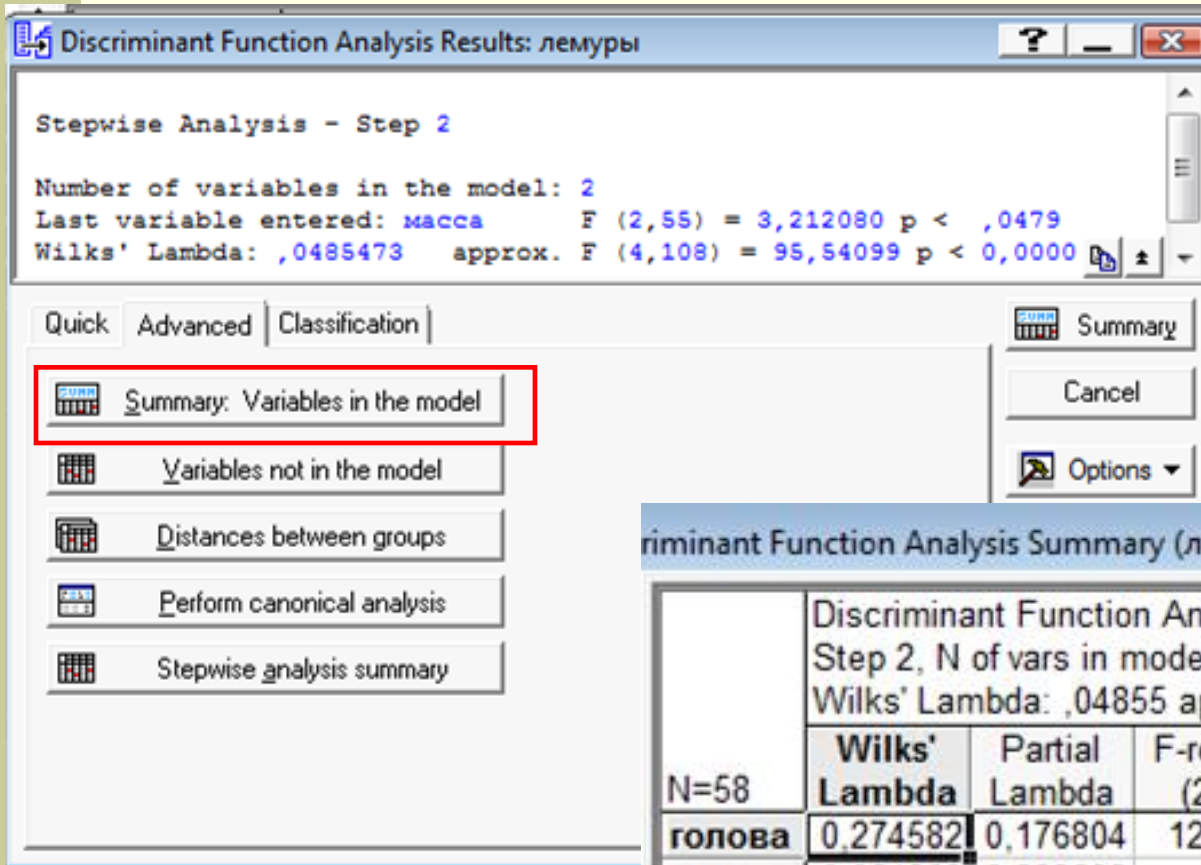




Variables currently not in the model (лемуры)						
Variables currently not in the model (лемуры)						
Df for all F-tests: 2,55						
N=58	Wilks' Lambda	Partial Lambda	F to enter	p-level	Toler.	1-Toler. (R-Sqr.)
масса	0,274582	0,274582	72,6522	0,000000	1,000000	0,00
голова	0,054323	0,054323	478,7336	0,000000	1,000000	0,00
зуб верхний	0,578130	0,578130	20,0671	0,000000	1,000000	0,00

Wilk's lambda – статистика, оценивает мощность дискриминации модели после введения в неё переменной. Чем она меньше – тем больше вклад (доля необъяснённой изменчивости)

F to enter – статистика для оценки достоверности вклада переменной в дискриминацию.



Пройдём Шаг 1 и Шаг 2.
 Можно посмотреть, какие
 переменные уже включены в
 анализ.

Discriminant Function Analysis Summary (лемуры)						
Discriminant Function Analysis Summary (лемуры) Step 2, N of vars in model: 2; Grouping: вид (3 grps) Wilks' Lambda: ,04855 approx. F (4,108)=95,541 p<0,0000						
N=58	Wilks' Lambda	Partial Lambda	F-remove (2,54)	p-level	Toler.	1-Toler. (R-Sqr.)
голова	0,274582	0,176804	125,7112	0,000000	0,792212	0,207788
масса	0,054323	0,893682	3,2121	0,048078	0,792212	0,207788

Partial lambda - статистика для вклада переменной в дискриминацию между совокупностями. Чем она меньше, тем больше вклад переменной.

Переменная **Голова** лучше помогает различать виды, чем **Масса**.

Discriminant Function Analysis Results: лемуры

Stepwise Analysis - Step 3 (Final Step)

Number of variables in the model: 3
 Last variable entered: зуб верх F (2,53) = 1,581093 p < ,2153
 Wilks' Lambda: ,0458138 approx. F (6,106) = 64,87177 p < 0,0000

Quick | Advanced | Classification

Summary: Variables in the model
 Variables not in the model
 Distances between groups
 Perform canonical analysis
 Stepwise analysis summary

Summary
 Cancel
 Options

у видами

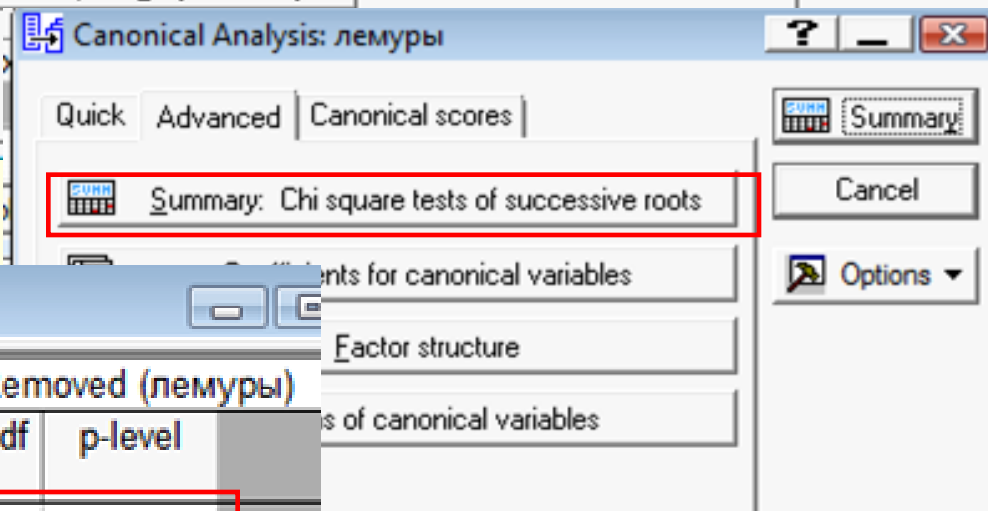
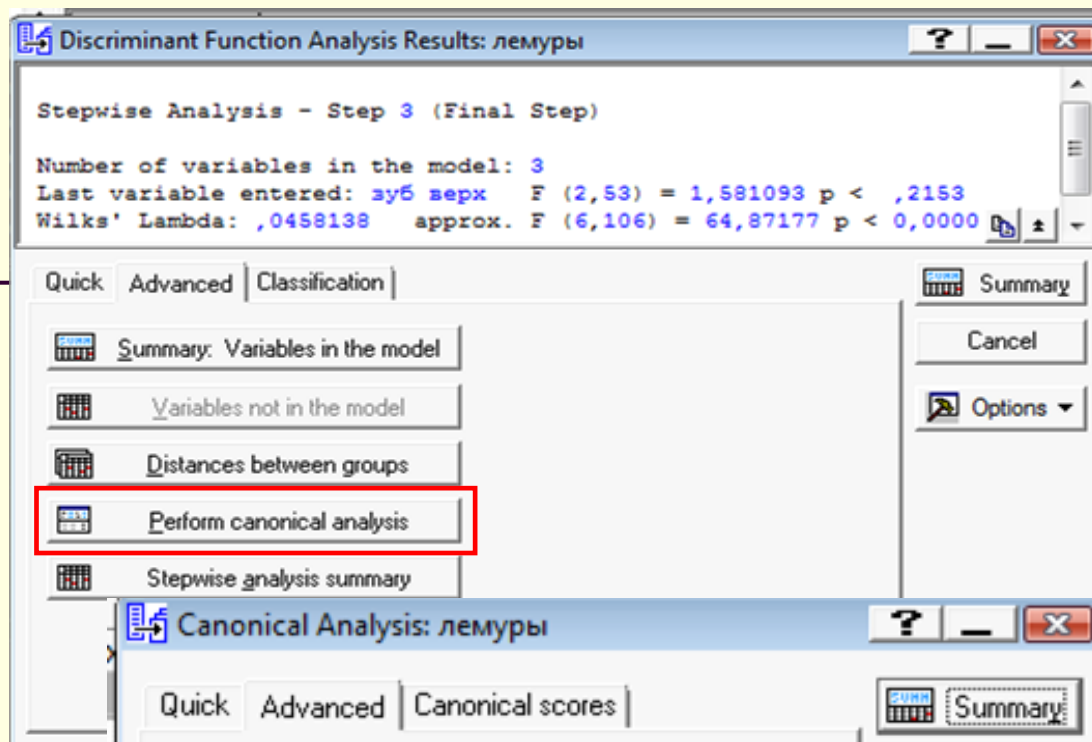
Discriminant Function Analysis Summary (лемуры)

Discriminant Function Analysis Summary (лемуры)						
Step 3, N of vars in model: 3; Grouping: вид (3 grps)						
Wilks' Lambda: ,04581 approx. F (6,106)=64,872 p<0,0000						
N=58	Wilks' Lambda	Partial Lambda	F-remove (2,53)	p-level	Toler.	1-Toler. (R-Sqr.)
голова	0,257081	0,178208	122,2028	0,000000	0,775488	0,224512
масса	0,051783	0,884736	3,4524	0,038955	0,728760	0,271241
зуб верхний	0,048547	0,943696	1,5811	0,215300	0,919741	0,080259

Partial lambda: Переменная Голова даёт вклад больше всех, а вклад Зуба – незначительный.

Создание дискриминантной функции

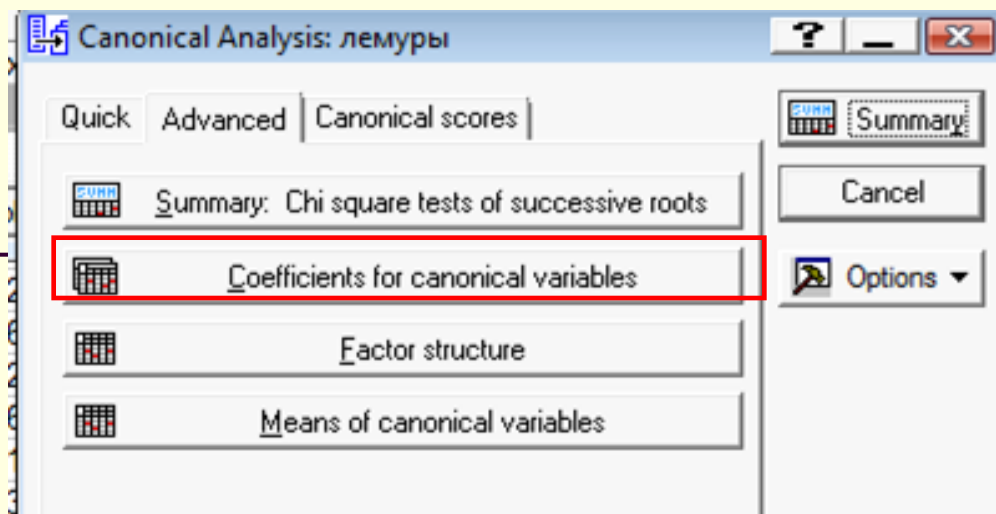
Дискриминантных функций -2



Square Tests with Successive Roots Removed (лемуры)

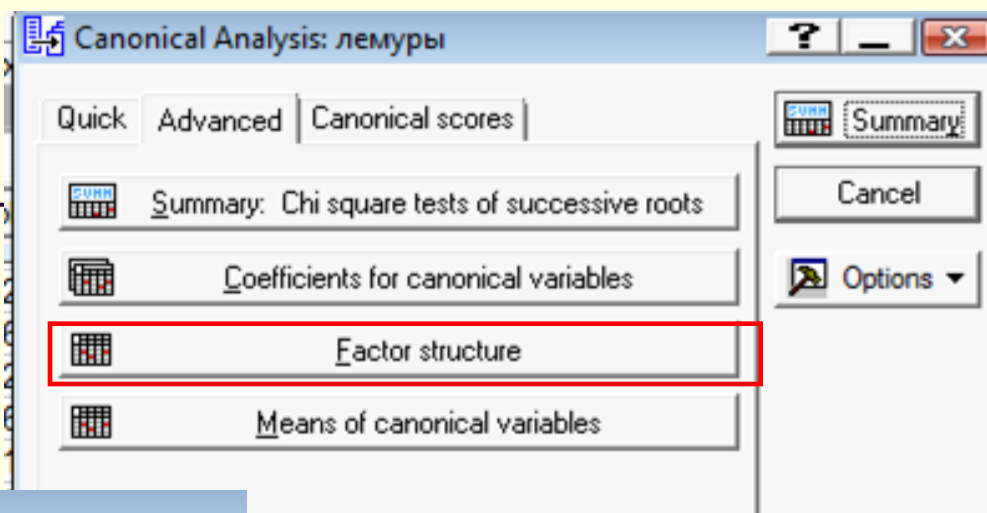
Chi-Square Tests with Successive Roots Removed (лемуры)						
Roots Removed	Eigen-value	Canonical R	Wilks' Lambda	Chi-Sqr.	df	p-level
0	18,62227	0,974186	0,045814	166,4911	6	0,000000
1	0,11238	0,317850	0,898972	5,7512	2	0,056382

Какой вклад внесли переменные в различение групп дискриминантными функциями?



Standardized Coefficients (лемуры)				
Variable	Standardized Coefficients (лемуры) for Canonical Variables			
	Root 1	Root 2		
голова	-1,04774	0,42091		
масса	0,17011	-1,13742		
зуб верхний	-0,25299	0,06861		
Eigenval	18,62227	0,11238		
Cum.Prop	0,99400	1,00000		

Standardized coefficients – коэффициенты для сравнения значимости (eigenvector). **Голова** лучше всех позволяет различать группы



Factor Structure Matrix (лемуры)

Factor Structure Matrix (лемуры) Correlations Variables - Canonical Roots (Pooled-within-groups correlations)				
Variable	Root 1	Root 2		
голова	-0,966831	-0,098483		
масса	-0,369679	-0,928690		
зуб верхний	-0,197236	-0,216579		

Canonical Analysis: лемуры

Quick | Advanced | Canonical scores

Summary: Chi square tests of successive roots

Coefficients for canonical variables

Factor structure

Means of canonical variables

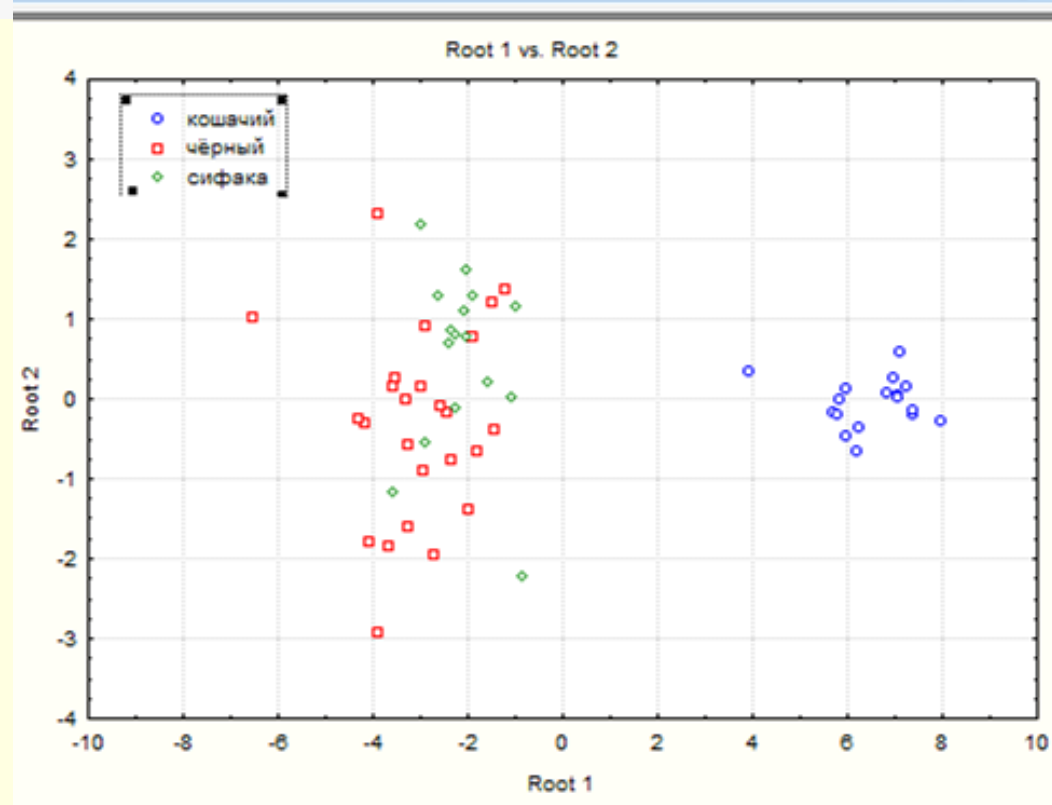
Summary

Cancel

Options

Means of Canonical Variables (лемуры)

	Root 1	Root 2
Group		
кошачий	6,49871	-0,046369
чёрный	-3,06030	-0,290091
сифака	-2,12316	0,502534



Классификация

Discriminant Function Analysis Results: лемуры

Stepwise Analysis - Step 3 (Final Step)

Number of variables in the model: 3
 Last variable entered: зуб верх F (2,53) = 1,581093 p < ,2153
 Wilks' Lambda: ,0458138 approx. F (6,106) = 64,87177 p < 0,0000

Quick | Advanced | Classification

Classification functions

A priori classification probabilities

- ☒ Proportional to group sizes
- ☐ Same for all groups
- ☐ User defined

Score to save for each case

- ☒ Save classification for case
- ☐ Save distance for case
- ☐ Save posterior probability for case

Max. number of cases in a single results spreadsheet: 100000

Classification Functions; grouping: вид (лемуры)

Variable	кошачий p=,29310	чёрный p=,43103	сифака p=,27586
голова	9,068	12,433	12,213
масса	-0,024	-0,026	-0,028
зуб верхний	53,984	61,209	60,659
Constant	-382,876	-662,947	-636,011

Функции классификации: получаем для них коэффициенты и можем классифицировать новых лемуров (взять новую особь, посчитать для неё функцию для каждой группы и отнести в ту группу, для которой значение будет наибольшим)

Значения p – вероятности случайного причисления лемура к той или иной группе, исходя из размеров группы.

Discriminant Function Analysis Results: лемуры

Stepwise Analysis - Step 3 (Final Step)

Number of variables in the model: 3
 Last variable entered: **зуб перх** $F(2,53) = 1,581093$ $p < ,2153$
 Wilks' Lambda: **,0458138** approx. $F(6,106) = 64,87177$ $p < 0,0000$

Quick | Advanced | Classification

Classification functions
 Use selection conditions to classify selected cases only [SELECT CASES](#) [Select](#)

Classification matrix

Classification of cases

A priori classification probabilities
☒ Proportional to group sizes
☐ Same for all groups
☐ User defined

Score to save for each case
☒ Save classification for case
☐ Save distance for case
☐ Save posterior probability for case

Max. number of cases in a single results spreadsheet: 100000

[Summary](#) [Cancel](#) [Options](#)

Classification Matrix (лемуры)

Classification Matrix (лемуры)				
Rows: Observed classifications				
Columns: Predicted classifications				
Group	Percent Correct	кошачий p=,29310	чёрный p=,43103	сифака p=,27586
кошачий	100,0000	17	0	0
чёрный	80,0000	0	20	5
сифака	75,0000	0	4	12
Total	84,4828	17	24	17

- Squared Mahalanobis Distances from Group Centroids (лемуры...)

Case	Observed Classif.	кошачий p=.29310	чёрный p=.43103	сифака p=.27586
1	кошачий	1,1964	106,8041	88,3084
2	чёрный	95,9080	0,1279	2,4912
3	сифака	79,7296	1,4777	0,1531
4	кошачий	0,5235	85,6465	70,4008
* 5	чёрный	63,2996	2,8465	1,4498
6	кошачий	1,3327	109,4110	91,2075
* 7	сифака	77,1818	0,7576	0,5151
8	чёрный	98,0735	1,9369	5,9141
9	чёрный	107,3656	3,1768	8,2933
10	сифака	95,3058	6,1445	3,6353
* 11	сифака	106,7712	5,1732	9,0136
12	чёрный	83,3294	3,8958	4,0371
13	чёрный	90,2732	0,3758	2,6620
14	сифака	72,7316	4,4586	1,2447
15	чёрный	117,1906	8,4964	15,7356
16	чёрный	95,9827	0,1566	1,6663
17	сифака	87,6743	5,0135	3,2262
18	чёрный	89,9184	0,2069	0,8619

Discriminant Function Analysis Results: лемуры

Stepwise Analysis - Step 3 (Final Step)

Number of variables in the model: 3
 Last variable entered: зуб верх $F(2,53) = 1,581093$ $p < ,2153$
 Wilks' Lambda: ,0458138 approx. $F(6,106) = 64,87177$ $p < 0,0000$

Quick | Advanced | Classification | Summary

☒ Classification functions
 Use selection conditions to classify selected cases only [SELECT CRISIS](#) [Select](#)

☒ Classification matrix

☒ Classification of cases

☒ Squared Mahalanobis distances

☒ Posterior probabilities

☒ Save scores

A priori classification probabilities
☒ Proportional to group sizes
☐ Same for all groups
☐ User defined

Score to save for each case
☒ Save classification for case
☐ Save distance for case
☐ Save posterior probability for case

Max. number of cases in a single results spreadsheet: 100000

[Cancel](#) [Options](#)

Требования к выборкам для дискриминантного анализа

точности такие же, как для MANOVA



1. Многомерное нормальное распределение: довольно устойчив к отклонениям при **одинаковых размерах групп**, желательны одномерные нормальные распределения;
2. Очень чувствителен к **аутлаерам**
3. Еще более чувствителен к **гетерогенности** дисперсий (**необходимо** проверить гомогенность для отдельных переменных)
4. Чем больше переменных в анализе, тем чувствительнее модель к нарушениям этих требований.
5. Не должно быть чрезмерно коррелирующих друг с другом переменных.

Разбиение выборки на две:

- **Анализируемая выборка (analysis sample)**
- Часть общей выборки, которую используют для вычисления дискриминантной функции.
- **Проверочная выборка (validation sample)**
- Часть общей выборки, которую используют для проверки результатов расчета на основании анализируемой выборки.

Определение коэффициентов дискриминантной функции

- **Прямой метод (direct method)**
- в котором дискриминантную функцию вычисляют при одновременном введении всех предикторов.
- В этом случае учитывается каждая независимая переменная
- **Пошаговый дискриминантный анализ (stepwise discriminant analysis)**
- в котором предикторы вводятся последовательно, в зависимости от их способности различить группы.
- Этот метод лучше применять в ситуации, когда исследователь хочет отобрать подмножество предикторов для включения их в дискриминантную функцию.

Постановка задачи: отдых на курорте

- Семьям, **которые отдыхали на курорте** в последние два года, присвоен код 1; тем же, которые **не посетили курорт** за указанный период времени, присвоен код 2.
- Обе выборки (как анализируемая, так и проверочная) сбалансированы с точки зрения посещаемости курорта. **Анализируемая** выборка содержит 15 семей каждой категории, а **проверочная** — по 6 семей каждой категории.
- Кроме того, получены данные о **доходе, отношении к путешествию, значении, придаваемом семейному отдыху, размеру семьи (размер семьи) и возрасту главы семьи (возраст)**.

Семейный отдых

Малхотра диск анализ отдых - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

Label Values Missing

1	Номер	None	None
2	Посещение курорта	{1, отдых в ...	None
3	Ежегодный доход	None	None
4	Отношение к путешествию	None	None
5	Значение, придаваемое семейному отдыху	None	None
6	Размер семьи	None	None
7	Возраст главы семьи	None	None
8			

Data View Variable View

Малхотра диск анализ отдых - SPSS Data Editor

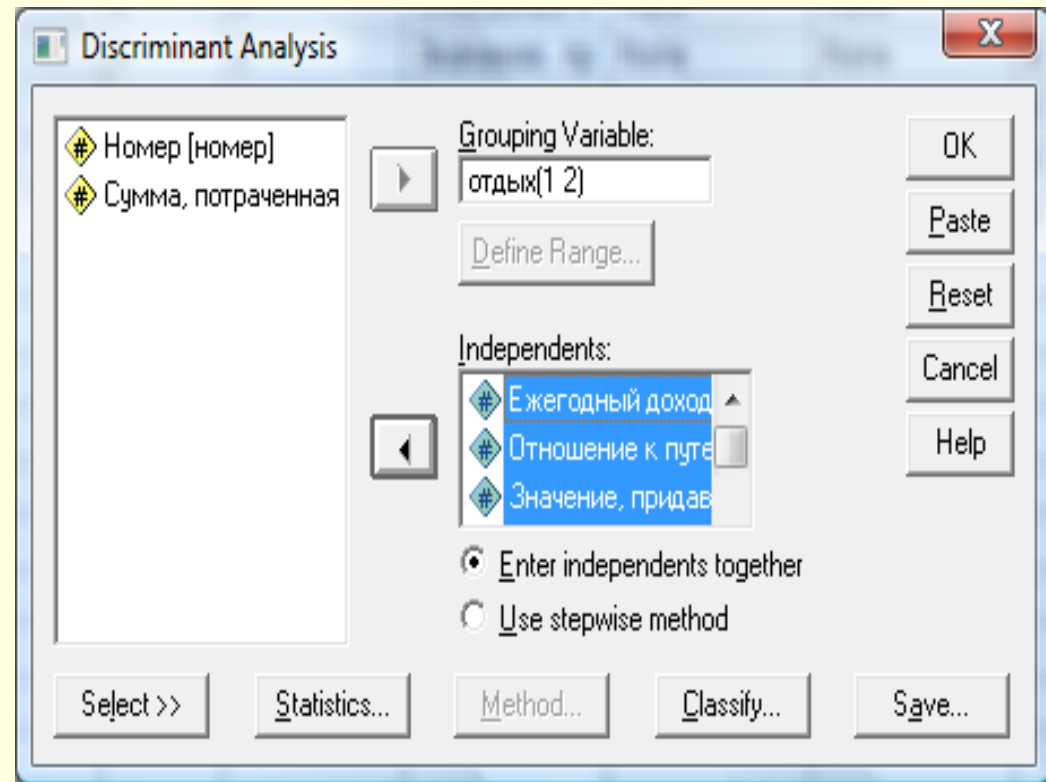
File Edit View Data Transform Analyze Graphs Utilities Window Help

1:

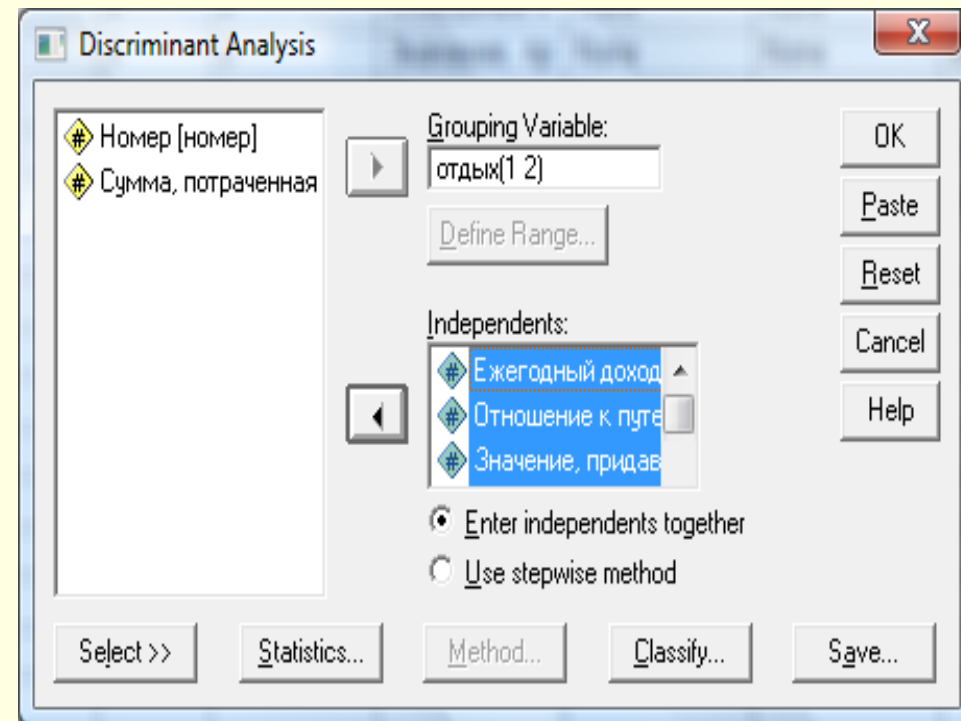
	номер	отдых	доход	путешест	значение	размер	возраст
1	1	1	50.20	5	8	3	43
2	2	1	70.30	5	7	4	61
3	3	1	52.90	7	5	6	52
4	4	1	46.50	7	5	5	36
5	5	1	52.70	6	6	6	55
6	6	1	75.00	8	7	5	68
7	7	1	46.20	5	3	3	62
8	8	1	57.00	2	4	6	51

Дискриминантный анализ в SPSS

Меню Analyze,
Classify,
Discriminant
Analysis



- Поле **Grouping Variable** (группирующая переменная) предназначено для задания единственной зависимой переменной
- В список **Independent** помещаются независимые переменные или предикторы
- **Enter independents together**
- **Use stepwise method** выбор варианта ДА





Кнопка “Statistics” и “Classify”

Discriminant Analysis: Statistics

Descriptives

- ☒ Means
- ☒ Univariate ANOVAs
- ☐ Box's M

Function Coefficients

- ☐ Fisher's
- ☒ Unstandardized

Matrices

- ☒ Within-groups correlation
- ☐ Within-groups covariance
- ☐ Separate-groups covariance
- ☐ Total covariance

Continue

Cancel

Help

Discriminant Analysis

Номер [номер]

Grouping Variable: отдых(1 2)

Define Range...

Independents:

- # Ежегодный доход
- # Отношение к путе
- # Значение, придав

☐ Enter independents together

☒ Use stepwise method

Select >> Statistics... Method... Classify... Save...

OK Paste Reset Cancel Help

Discriminant Analysis: Classification

Prior Probabilities

- ☒ All groups equal
- ☐ Compute from group sizes

Use Covariance Matrix

- ☒ Within-groups
- ☐ Separate-groups

Display

- ☒ Casewise results
 - ☐ Limit cases to first:
- ☒ Summary table
- ☐ Leave-one-out classification

Plots

- ☐ Combined-groups
- ☒ Separate-groups
- ☐ Territorial map

☐ Replace missing values with mean

Continue

Cancel

Help

Результаты: групповые статистики (1)

Group Statistics

Посещение курорта		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
1	Ежегодный доход	59.7200	10.16951	15	15.000
	Отношение к путешествию	5.3333	1.91485	15	15.000
	Значение, придаваемое семейному отдыху	5.8000	1.82052	15	15.000
	Размер семьи	4.4667	1.30201	15	15.000
	Возраст главы семьи	53.7333	8.77062	15	15.000
2	Ежегодный доход	41.8467	7.51588	15	15.000
	Отношение к путешествию	4.3333	1.95180	15	15.000
	Значение, придаваемое семейному отдыху	4.2000	1.89737	15	15.000
	Размер семьи	2.8667	.91548	15	15.000
	Возраст главы семьи	50.1333	8.27101	15	15.000
Total	Ежегодный доход	50.7833	12.64178	30	30.000
	Отношение к путешествию	4.8333	1.96668	30	30.000
	Значение, придаваемое семейному отдыху	5.0000	2.00000	30	30.000
	Размер семьи	3.6667	1.37297	30	30.000
	Возраст главы семьи	51.9333	8.57395	30	30.000

Тест на равенство средних групп (2)

Tests of Equality of Group Means

	Wilks' Lambda	F	df 1	df 2	Sig.
Ежегодный доход	.483	29.966	1	28	.000
Отношение к путешествию	.933	2.006	1	28	.168
Значение, придаваемое семейному отдыху	.834	5.554	1	28	.026
Размер семьи	.649	15.158	1	28	.001
Возраст главы семьи	.954	1.338	1	28	.257

Если значимость < 0.05 то
средние двух групп значимо
различаются

Коэффициент Уилкса

- коэффициент Уилкса для каждого предиктора — это отношение внутригрупповой суммы квадратов к общей сумме квадратов.
- Его значение варьирует от 0 до 1.
- Большое значение X (около 1) указывает на то, что средние групп не должны различаться.
- Малые значения (около 0) указывают на то, что средние групп различаются.

Объединенная внутригрупповая корреляционная матрица (3)

Pooled Within-Groups Matrices

	Ежегодный доход	Отношение к пу тешеств ию	Значение, придав аемое семе йному отдыху	Размер семе й	Возраст главы семе й
Correlation					
Ежегодный доход	1.000	.135	.112	.017	.010
Отношение к пу тешеств ию	.135	1.000	.089	-.044	-.213
Значение, придаваемое семе йному отдыху	.112	.089	1.000	.048	.045
Размер семе й	.017	-.044	.048	1.000	-.007
Возраст главы семе й	.010	-.213	.045	-.007	1.000

указывает на низкие коэффициенты корреляции между предикторами. Ее вычисляют усреднением отдельных корреляционных матриц для всех групп.

Собственные значения и Лямба Уилкса (4)

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.751 ^a	100.0	100.0	.798

a. First 1 canonical discriminant functions were used in the analysis.

Каноническая корреляция - коэффициент между рассчитанными значениями дискриминантной функции и показателем принадлежности к группе, чем он ближе к 1, тем лучше.

Собственное значение (eigenvalue) — это отношение межгрупповой суммы квадратов к внутригрупповой сумме квадратов (лучшие значения - больше 1)

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.364	25.803	5	.000

Лямба Уилкса показывает, значимо ли различаются в двух группах средние значения дискриминантной функции. Если значимость < 0,05 - значимое различие



Стандартизованные коэффициенты и структурная матрица (5)

Standardized Canonical Discriminant Function Coefficients

	Function
	1
Ежегодный доход	.727
Отношение к путешествию	.148
Значение, придаваемое семейному отдыху	.208
Размер семьи	.541
Возраст главы семьи	.184

Коэффициенты дискриминантных функций используют как множители для нормированных переменных, т.е. переменных с нулевым средним и дисперсией, равной 1.

Structure Matrix

	Function
	1
Ежегодный доход	.782
Размер семьи	.556
Значение, придаваемое семейному отдыху	.337
Отношение к путешествию	.202
Возраст главы семьи	.165

Ordered within-groups correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation within function.

Корреляционная матрица между предикторами и дискриминантной функцией (переменные расположены в соответствии с размером корреляции)

Нестандартизованные коэффициенты (6)

anonical Discriminant Function Coefficient

	Function
	1
Ежегодный доход	.081
Отношение к пу тешествию	.077
Значение, придав аемое смейному отдыху	.112
Размер семьи	.481
Возраст главы семьи	.022
(Constant)	-7.944

Unstandardized coefficients

(ненормированные)—
это коэффициенты
переменных, когда
они измерены в
первоначальных
единицах.

Дискриминантные функции, оцененные по центроидам

Functions at Group Centroids

	Function
Посещение ку рорта	1
1	1.278
2	-1.278

Unstandardized canonical discriminant functions evaluated at group means

это средние значения для дискриминантных показателей конкретной группы. Центроидов столько, сколько групп, т.е. один центроид для каждой группы

Данные о фактической и прогнозируемой группе и значения дискриминантов для каждого объекта

Casewise Statistics

Case Number	Actual Group	Highest Group						Second Highest Group			Discriminant Scores
		Predicted Group	P(D>d G=g)		P(G=g D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g D=d)	Squared Mahalanobis Distance to Centroid	Function 1	
			p	df							
Original	1	2**	.287	1	.633	1.136	1	.367	2.223	-.213	
	2	1	.367	1	.996	.812	2	.004	11.957	2.180	
	3	1	.854	1	.977	.034	2	.023	7.511	1.462	
	4	1	.245	1	.573	1.352	2	.427	1.943	.116	
	5	1	.789	1	.981	.071	2	.019	7.975	1.546	
	6	1	.032	1	1.000	4.604	2	.000	22.113	3.424	
	7	2**	.554	1	.853	.349	1	.147	3.863	-.687	
	8	1	1.000	1	.963	.000	2	.037	6.538	1.279	
	9	1	.810	1	.980	.058	2	.020	7.825	1.519	
	10	1	.368	1	.996	.810	2	.004	11.949	2.178	
	11	1	.179	1	.999	1.808	2	.001	15.220	2.623	
	12	1	.227	1	.998	1.459	2	.002	14.172	2.486	
	13	1	.711	1	.985	.137	2	.015	8.567	1.649	
	14	2**	.599	1	.873	.276	1	.127	4.126	-.753	
	15	1	.362	1	.718	.832	2	.282	2.704	.366	
	16	2	.596	1	.990	.281	1	.010	9.529	-1.809	
	17	2	.352	1	.996	.867	1	.004	12.163	-2.209	
	18	2	.804	1	.980	.062	1	.020	7.868	-1.527	
	19	2	.445	1	.789	.582	1	.211	3.217	-.515	
	20	2	.688	1	.904	.161	1	.096	4.646	-.877	
	21	2	.626	1	.989	.237	1	.011	9.264	-1.765	
	22	2	.435	1	.781	.609	1	.219	3.155	-.498	
	23	2								-.944	
	24	2								-.405	
	25	2								-.886	
	26	2								-1.174	
	27	2								-.620	
	28	2								-1.833	
	29	2								-1.962	
	30	2	.383	1	.996	.762	1	.004	11.763	-2.151	

Для объектов, отмеченных **
предсказанная и фактическая группа не
совпали

** - Misclassified case

Расшифровка терминов

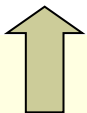
- **Actual group** (Фактическая группа)
- **Predicate group** (Предсказанная группа) - вычисленная для объекта с помощью дискр. функции.
- **Ungrouped** (Несгруппированный объект) - заранее неизвестна принадлежность к группе.
- **Discriminant scores** (Значения дискриминантной функции) – получаются при подстановке значений переменных для объекта в уравнение дискр. функции.

Результаты классификации (8)

Classification Results^a

			Predicted Group Membership		Total
			1	2	
Original	Count	Посещение курорта			
		1	12	3	15
		2	0	15	15
	%	1	80.0	20.0	100.0
		2	.0	100.0	100.0

a. 90.0% of original grouped cases correctly classified.



Показывает % верно классифицированных случаев. Три случая из 30 были неправильно классифицированы.