

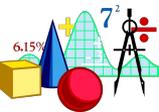
ДИСКРИМИНАНТНЫЙ АНАЛИЗ

**Стат. методы в
психологии
(Радчикова Н.П.)**



Цели

- В каких случаях применяется дискриминантный анализ
- Как применить дискриминантный анализ
- Как интерпретировать результаты дискриминантного анализа





Выбор метода прогнозирования

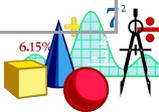
Вид зависимости	Зависимая переменная	Независимые переменные	Метод
Линейная	Шкала интервалов или равных отношений	1 переменная (шкала интервалов или равных отношений)	Простая линейная регрессия
	Шкала интервалов или равных отношений	Несколько любых переменных*	Множественная линейная регрессия
	Шкала порядка или наименований	Несколько любых переменных*	Дискриминантный анализ





Выбор метода прогнозирования

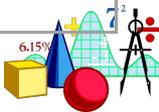
Вид зависимости	Зависимая переменная	Независимые переменные	Метод
Линейная	Шкала интервалов или равных отношений	1 переменная (шкала интервалов или равных отношений)	Простая линейная регрессия
	Шкала интервалов или равных отношений	Несколько любых переменных*	Множественная линейная регрессия
	Шкала порядка или наименований	Несколько любых переменных*	Дискриминантный анализ





Выбор метода прогнозирования

Вид зависимости	Зависимая переменная	Независимые переменные	Метод
Линейная	Шкала интервалов или равных отношений	1 переменная (шкала интервалов или равных отношений)	Простая линейная регрессия
	Шкала интервалов или равных отношений	Несколько любых переменных*	Множественная линейная регрессия
	Шкала порядка или наименований	Несколько любых переменных*	Дискриминантный анализ

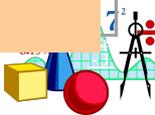




Выбор метода прогнозирования

Вид зависимости	Зависимая переменная	Независимые переменные	Метод
Линейная	Шкала интервалов или равных отношений	1 переменная (шкала интервалов или равных отношений)	Простая линейная регрессия
	Шкала интервалов или равных отношений	Несколько любых переменных*	Множественная линейная регрессия
	Шкала порядка или наименований	Несколько любых переменных*	

**Дискри-
минантный
анализ**



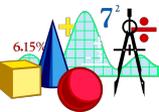


Шкалы наименований

**Мы уже знаем, что можно использовать
дихотомические шкалы.**

**А что делать, если попалась шкала
наименований?**

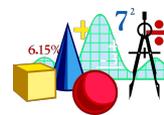
**Не спешите расстраиваться! Надо ее
просто перекодировать!**





Дискриминантный анализ

Альтернатива множественного регрессионного анализа для случая, когда зависимая переменная качественная (категориальная).



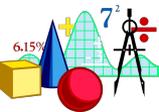


Дискриминантный анализ

Основная цель:

Выявление структуры исследуемого
множества объектов

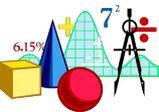
(структура – набор основных
факторов (шкал), по которым
различаются и могут быть описаны
объекты)





Основная задача

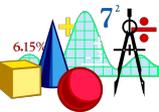
По значениям дискриминантных переменных для объектов получить значения классифицирующей переменной, то есть определить классы, в которые попадают эти объекты.





Основная задача

На основании некоторых признаков (независимых переменных) объект или индивидуум может быть причислен к одной из двух (или к одной из нескольких) **заранее** заданных групп.





Ограничения

В случае дискриминантного анализа предполагается, что

- зависимая переменная одна и представлена в шкале наименований**
- независимых переменных несколько**





Представление данных

группа	X1	X2	...	X34
Гр1	23	2,5		123
Гр1	21	1,7		131
...
Гр2	24	1,5		148
Гр2	21	2,1		133
...

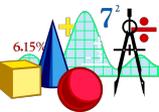




Основная идея

Дискриминантная функция

$$z = b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \dots + b_n \cdot x_n + b_0$$

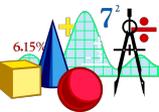




Основная идея

Наша цель:

Определить коэффициенты b ,
чтобы по значениям
дискриминантной функции можно
было с максимальной четкостью
провести разделение по группам.





Пример для двух групп

группа	X1	X2	...	X34
Гр1	23	2,5		123
Гр1	21	1,7		131
...
Гр2	24	1,5		148
Гр2	21	2,1		133
...





Основная идея

Строим дискриминантную функцию

$$z = b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \dots + b_n \cdot x_n + b_0,$$

такую, что разница между средними значениями \bar{z}_1 и \bar{z}_2 , полученными на множествах значений НП для разных групп максимальна.





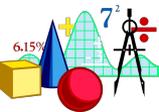
Основная идея

$|\bar{z}_1 - \bar{z}_2| \Rightarrow$ максимум

Фишер показал, что

$$b = S^{-1} \cdot (\bar{x}_1 - \bar{x}_2),$$

где S – ковариационная матрица

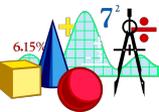
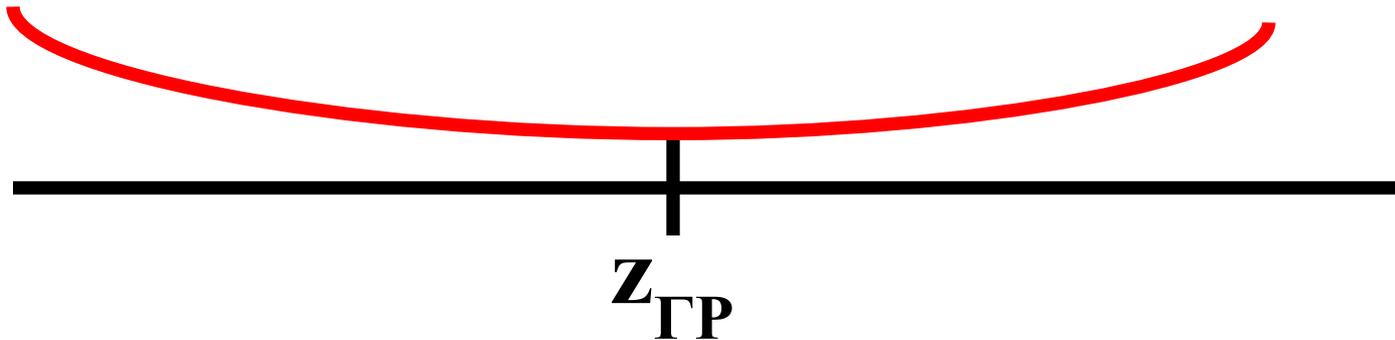




Основная идея

Классификация происходит посредством определения величины $z_{ГР}$

$$z_{ГР} = (\bar{z}_1 - \bar{z}_2) / 2$$





Основная идея

Предположив, что z_1 – большее из двух средних, получаем правило:

Случай относится к группе 1, если $z_i - z_{ГР} > 0$

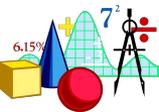
Случай относится к группе 2, если $z_i - z_{ГР} \leq 0$





Основная идея

\bar{z}_1 и \bar{z}_2 называются центроидами
групп





Пример для двух групп

Данные

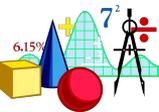
GENDER – пол испытуемого;

EDUC – образование испытуемого

(количество лет, которые бедняга потратил на учебу);

JCAT – вид профессиональной деятельности (1 – клерк, 2- охранник, 3 – менеджер);

SALARY – зарплата в настоящий момент;





Пример для двух групп

Данные

SAL_BEG – начальная зарплата на этой работе;

JTIME – трудовой стаж на данном рабочем месте (число месяцев);

PREVEX – предыдущий опыт – стаж до поступления на данную работу;

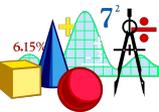
MINORITY – принадлежит ли испытуемый к национальному меньшинству (0 – нет, 1 – да).





Пример для двух групп

**Попробуем предсказать,
принадлежит ли человек к
национальному меньшинству на
основании его зарплаты и
образования**





Пример для двух групп

- **Что мы получим в результате применения дискриминантного анализа?**
- **Как это интерпретировать?**





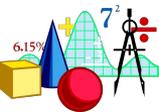
Модуль дискриминантного анализа

Discriminant Analysis

Statistics \Rightarrow

Multivariate Exploratory Techniques \Rightarrow

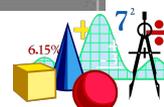
Discriminant Analysis



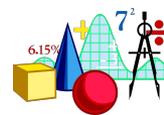
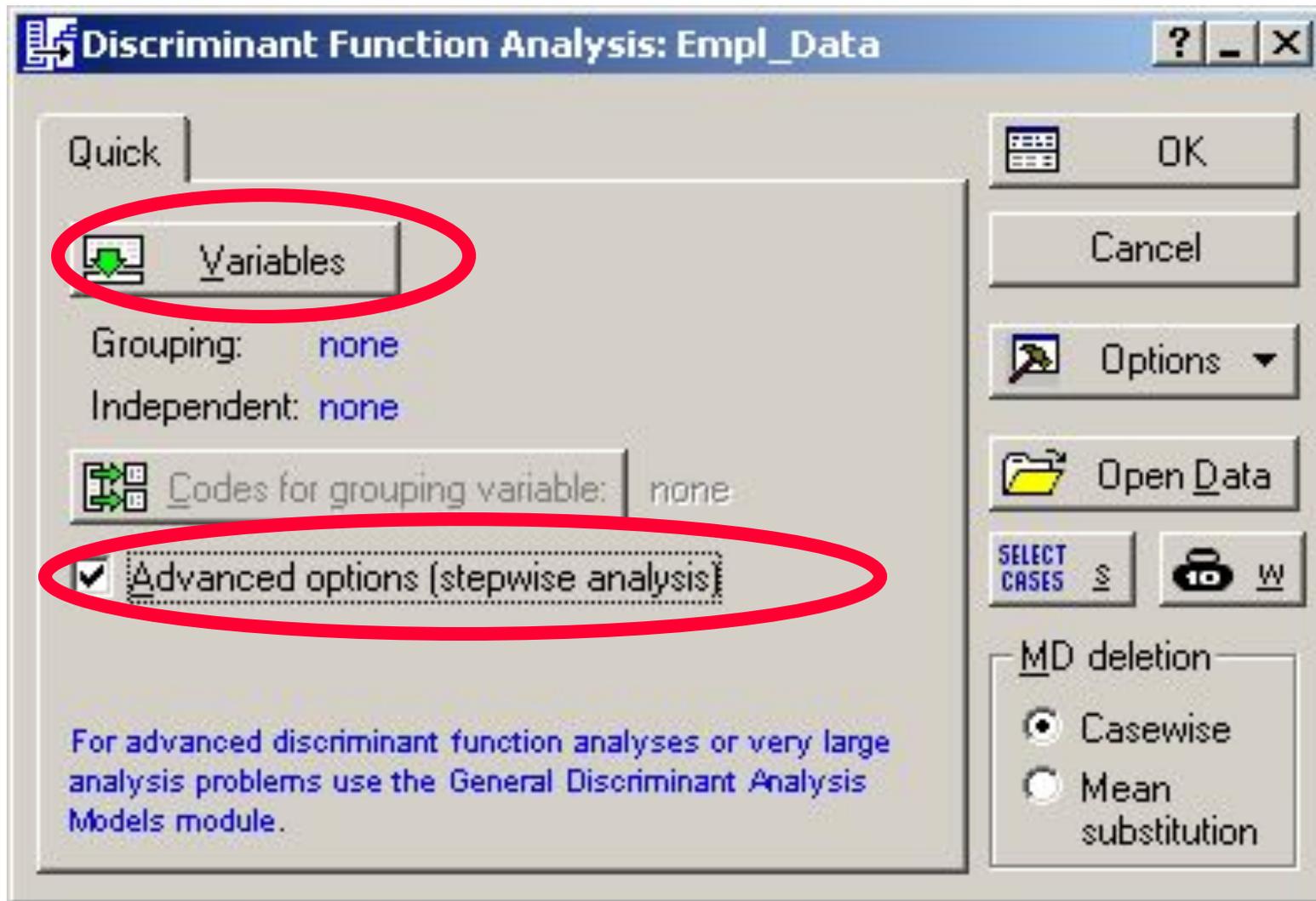
Модуль дискриминантного анализа

The screenshot displays the STATISTICA software interface with the 'Statistics' menu open. The 'Discriminant Analysis' option is highlighted with a red circle. The background shows a data table with columns for ID, GENDER, and other variables.

	1	2							
	ID	GENDER							
1	1	M							
2	2	M							
3	3	Ж							
4	4	Ж							
5	5	M							
6	6	M							
7	7	M							
8	8	Ж							
9	9	Ж							
10	10	Ж							
11	11	Ж							
12	12	M							
13	13	M							
14	14	Ж							
15	15	M							
16	16	M	12	1	40,800	15,000	97	66	0
17	17	M	15	1	46,000	14,250	97	24	0
18	18	M	16	3	103,750	27,510	97	48	0
19	19	M	16	3	103,750	27,510	97	70	0
20	20	Ж	12	1	42,300	14,250	97	103	0
21	21	Ж	12	1	26,250	11,550	97	48	0
22	22	Ж	16	1	38,850	15,000	97	17	0

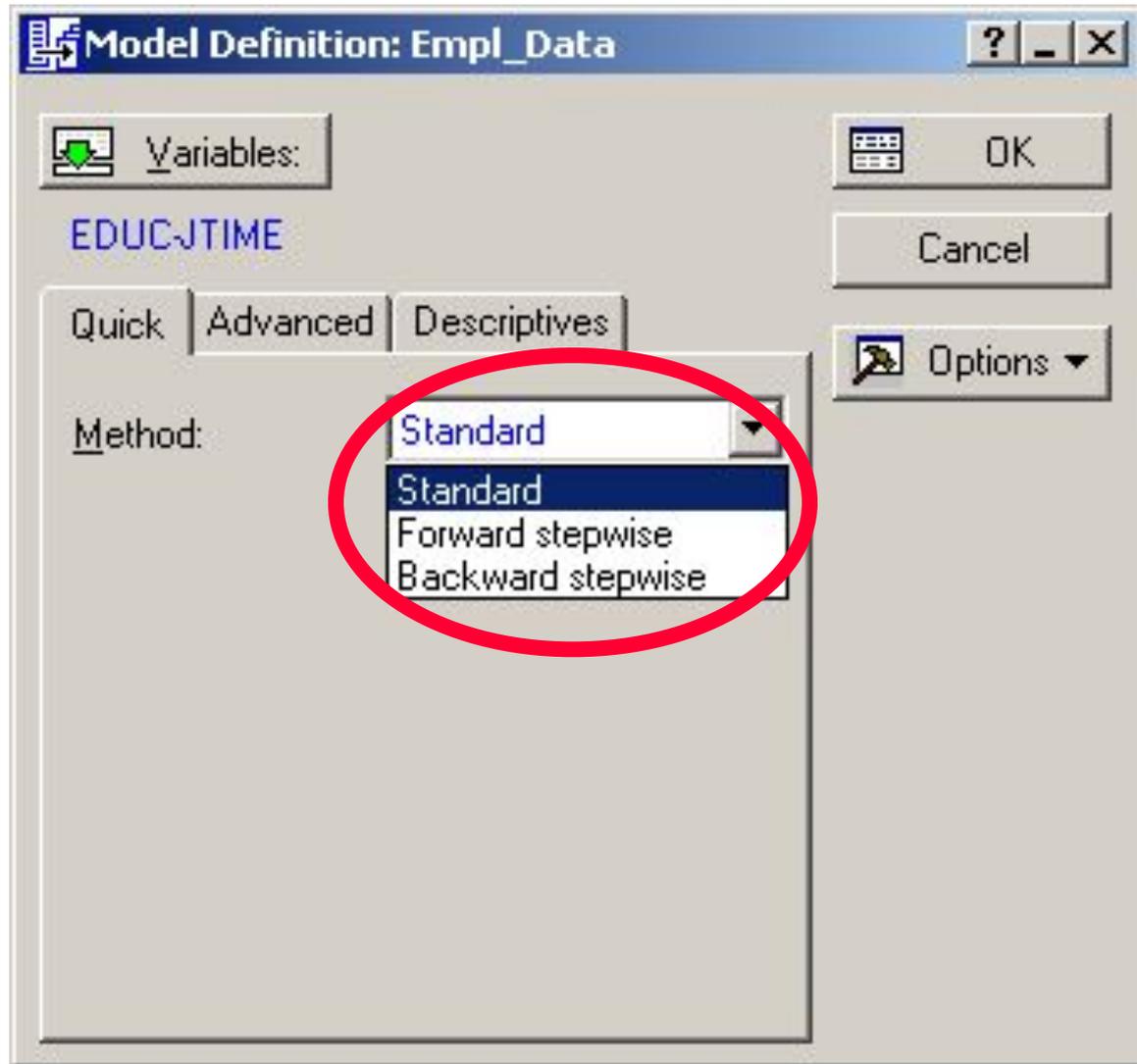


Модуль дискриминантного анализа





Модуль дискриминантного анализа



Получаем результаты (Quick):

Discriminant Function Analysis Results: Empl_Data

Number of variables in the model: 3

Wilks' Lambda: ,9681115 approx. F (3,470) =

Quick | Advanced | Classification

Summary: Variables in the model

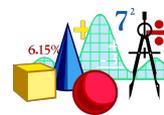
Variables not in the model

Cancel

Options ▾

By Group

**Анализ
переменных,
использующихся
в модели**

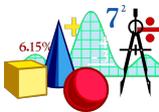




Variables in the Model:

DISCRIM. ANALYSIS	No. of vars in model	
	Wilks' Lambda: ,9678	
N=474	Wilks' Lambda	Partial Lambda
EDUC	,968114	,999732
SALARY	,973334	,994370
SAL_BEG	,968442	,999394

Лямбда Уилкса для модели с исключенной данной переменной. Изменяется от 0 (совершенное различие) до 1 (никакого различия)





Variables in the Model:

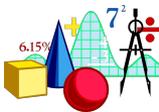
Discriminant Function Analysis Summary (emp)

DISCRIM. ANALYSIS No. of vars in model: 3;
Wilks' Lambda: ,96785 ap

N=474

	Wilks' Lambda	Partial Lambda	F-
EDUC	,968114	,999732	,1
SALARY	,973334	,994370	2,8
SAL_BEG	,968442	,999394	,2

Эта лямбда связана с вкладом данной переменной в различительную силу модели





Variables in the Model:

Discriminant Function Analysis Summary (empl_data.sta)

DISCRIM. No. of vars in model: 3; Grouping: MINORITY (2 grps)
ANALYSIS Wilks' Lambda: ,96785 approx. F (3,470)=5,2033 p< ,0015

N=474	Wilks' Lambda	Partial Lambda	F-remove (1,470)	p-level	Toler.	1-Toler. (R-Sqr.)
			,126120	,722649	,544077	,455923
			,660899	,103512	,337718	,662282
			,284960	,593721	,355688	,644312

Статистика дисперсионного анализа, показывающая вклад данной переменной в общее «дело» различения групп.





Variables in the Model:

Discriminant Function Analysis Summary (empl_data.sta)

DISCRIM. ANALYSIS No. of vars in model: 3; Grouping: MINORITY (2 grps)
Wilk's Lambda = .344960, F (3, 470) = 5,2023, p = .0015

N=474

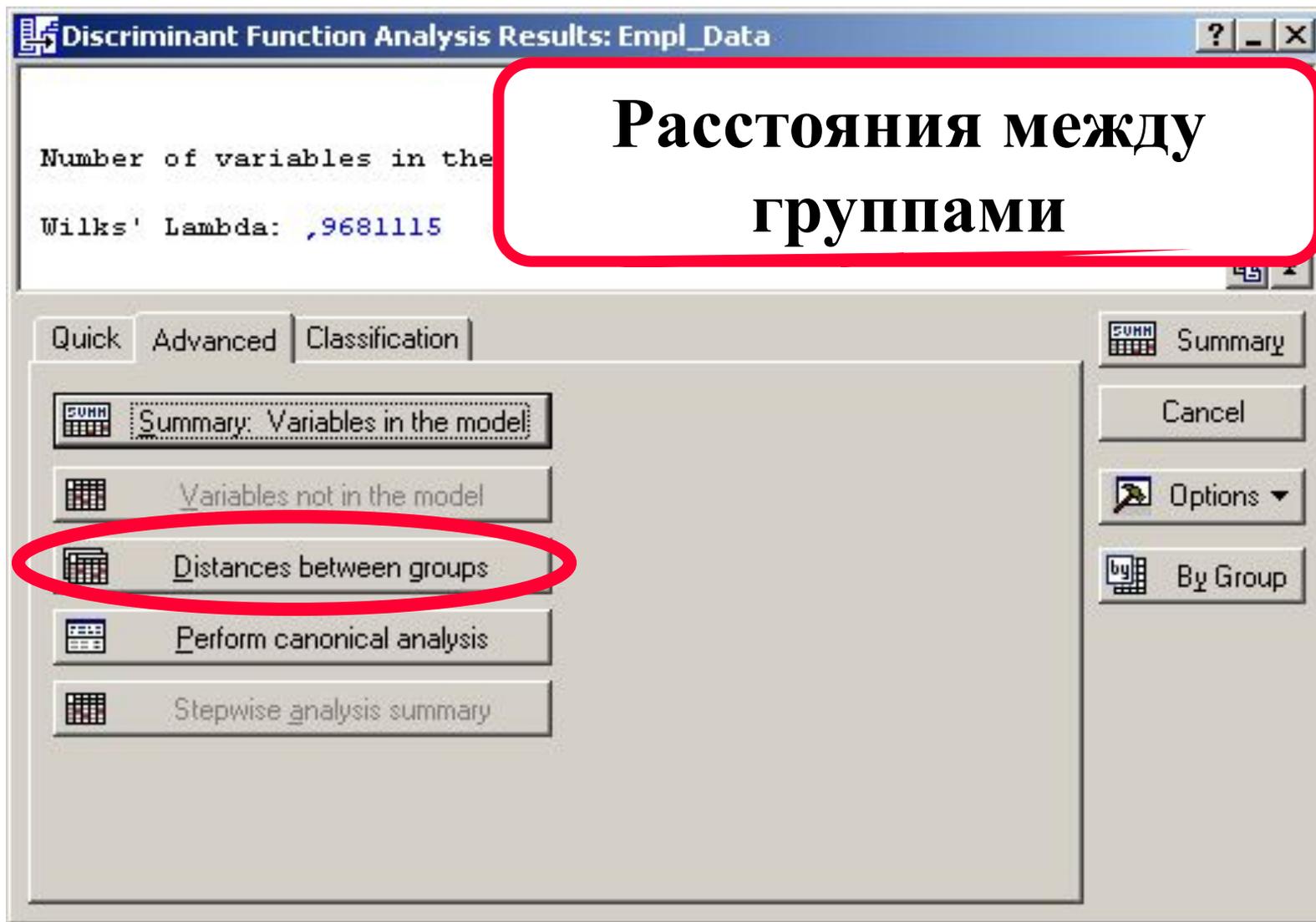
Variable	p-level	Toler.	1-Toler. (R-Sqr.)
EDUC	,722649	,544077	,455923
SALARY	,103512	,337718	,662282
SAL_BEG	,84960	,593721	,644312

Толерантность
– измеряет
избыточность
данной
переменной.

Толерантность 0,34 означает, что переменная на 66% объясняет то, что и другие переменные модели



Результаты анализа (Advanced)



The screenshot shows the 'Discriminant Function Analysis Results: Empl_Data' dialog box. The 'Advanced' tab is selected. The 'Distances between groups' button is circled in red. The 'Summary: Variables in the model' button is also highlighted with a red box. The Wilks' Lambda value is .9681115.

Discriminant Function Analysis Results: Empl_Data

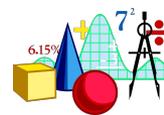
Number of variables in the model: 4
Wilks' Lambda: .9681115

Quick | **Advanced** | Classification

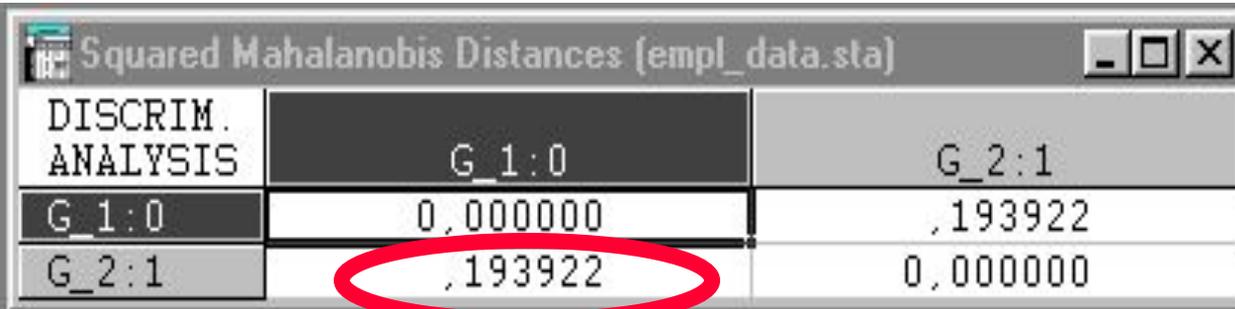
Summary: Variables in the model
Variables not in the model
Distances between groups
Perform canonical analysis
Stepwise analysis summary

Summary
Cancel
Options
By Group

Расстояния между
группами



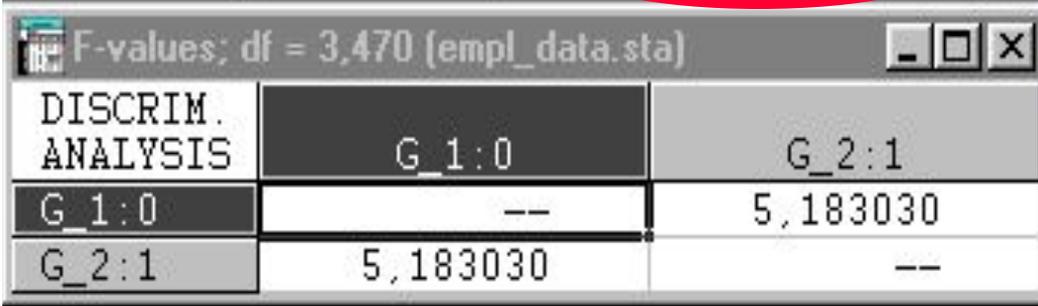
Distances between groups



Squared Mahalanobis Distances (empl_data.sta)

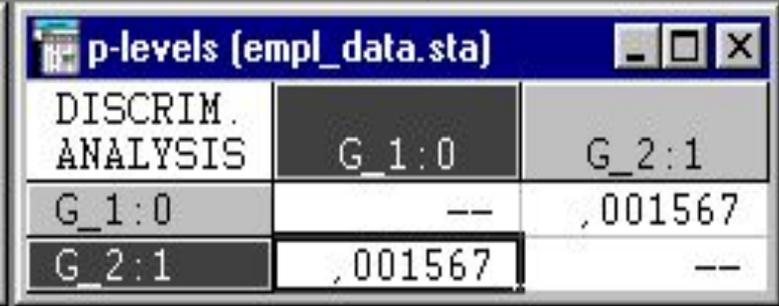
DISCRIM. ANALYSIS	G_1:0	G_2:1
G_1:0	0,000000	,193922
G_2:1	,193922	0,000000

The value .193922 in the bottom-left cell is circled in red.



F-values; df = 3,470 (empl_data.sta)

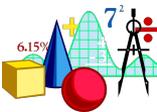
DISCRIM. ANALYSIS	G_1:0	G_2:1
G_1:0	--	5,183030
G_2:1	5,183030	--



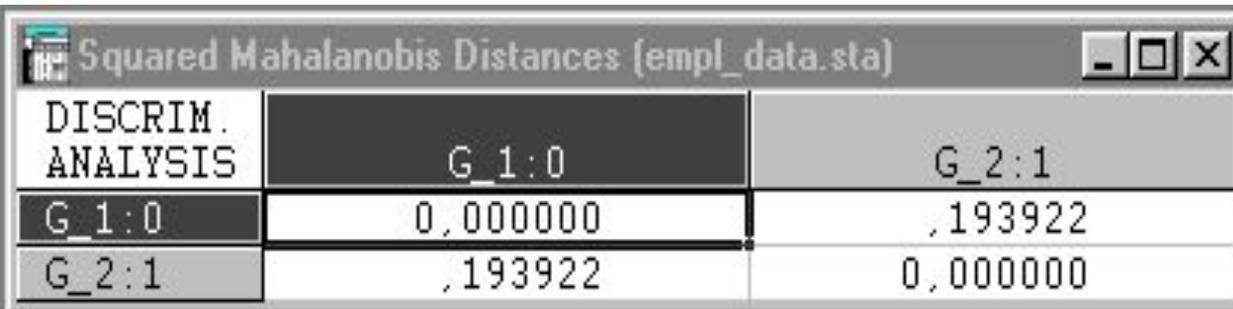
p-levels (empl_data.sta)

DISCRIM. ANALYSIS	G_1:0	G_2:1
G_1:0	--	,001567
G_2:1	,001567	--

**Расстояние Махаланобиса
между группами**

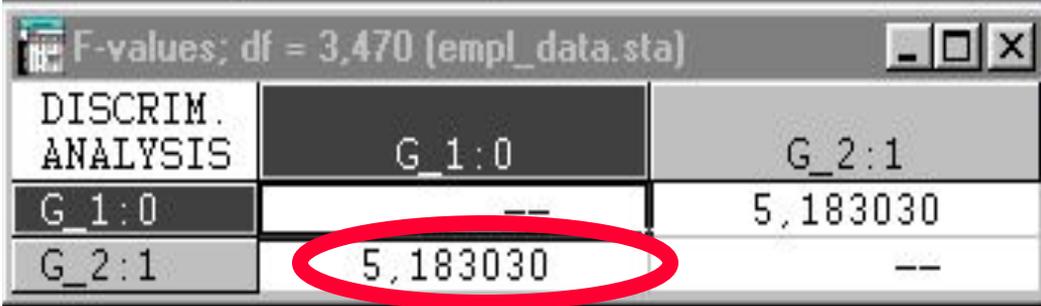


Distances between groups



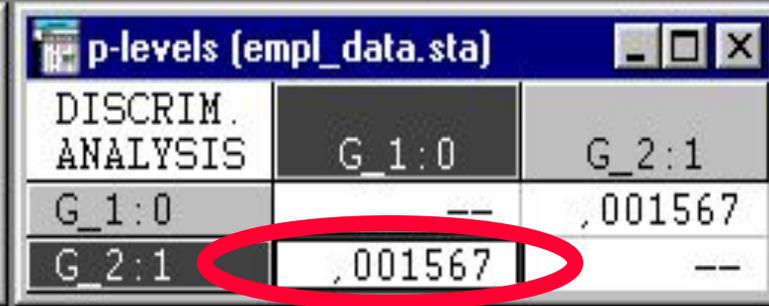
Squared Mahalanobis Distances (empl_data.sta)

DISCRIM. ANALYSIS	G_1:0	G_2:1
G_1:0	0,000000	,193922
G_2:1	,193922	0,000000



F-values; df = 3,470 (empl_data.sta)

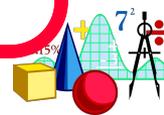
DISCRIM. ANALYSIS	G_1:0	G_2:1
G_1:0	--	5,183030
G_2:1	5,183030	--



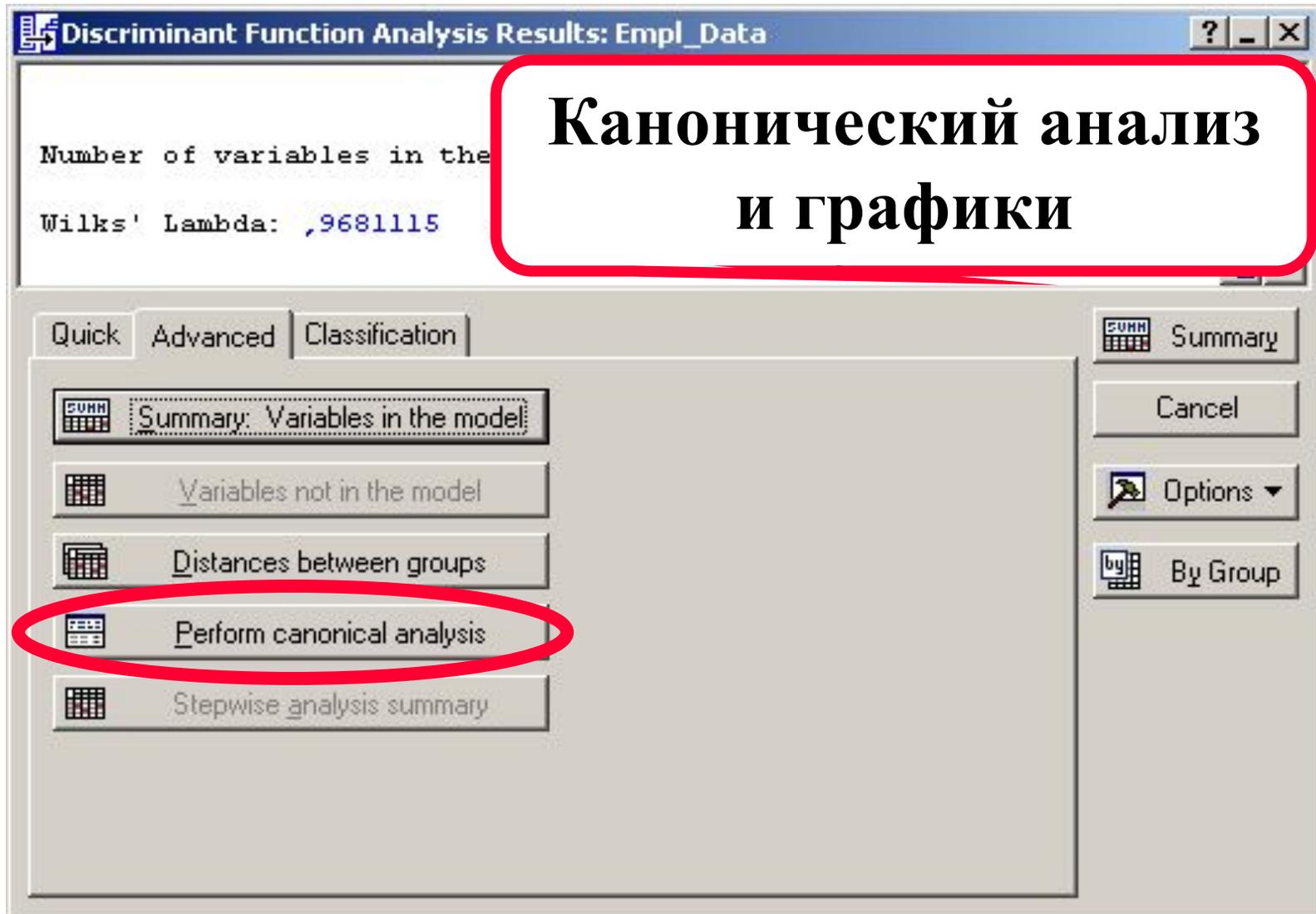
p-levels (empl_data.sta)

DISCRIM. ANALYSIS	G_1:0	G_2:1
G_1:0	--	,001567
G_2:1	,001567	--

Значение дисперсионного анализа и соответствующий уровень значимости для оценки расстояния между группами



Результаты анализа (Advanced)



The screenshot shows the 'Discriminant Function Analysis Results: Empl_Data' dialog box in SPSS. The 'Advanced' tab is selected. The 'Number of variables in the model' is set to 1, and 'Wilks' Lambda' is .9681115. The 'Perform canonical analysis' button is circled in red. Other buttons include 'Summary', 'Cancel', 'Options', and 'By Group'.

Discriminant Function Analysis Results: Empl_Data

Number of variables in the model: 1
Wilks' Lambda: .9681115

Quick | **Advanced** | Classification

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

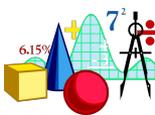
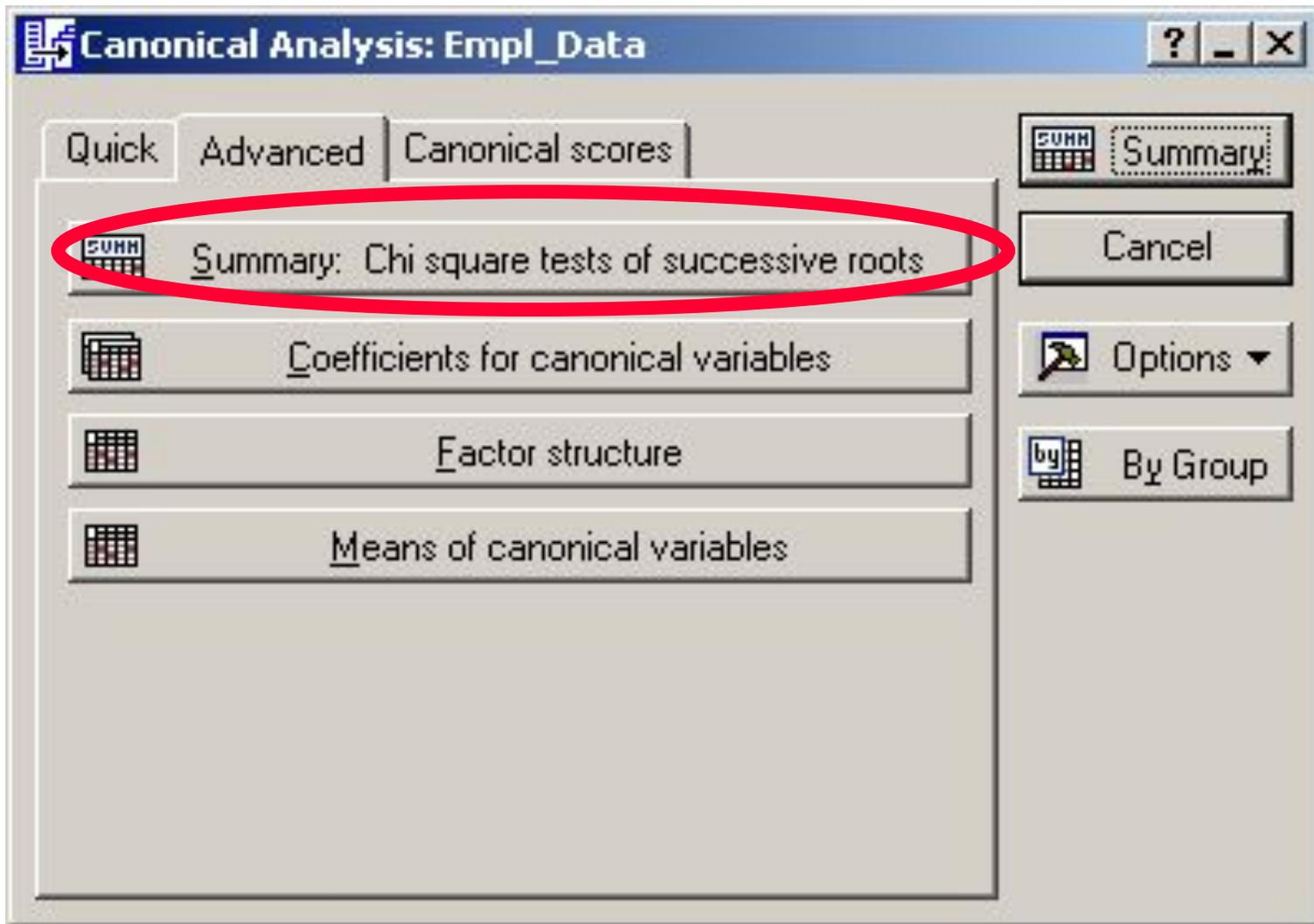
Summary
Cancel
Options
By Group

**Канонический анализ
и графики**





Canonical Analysis:





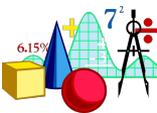
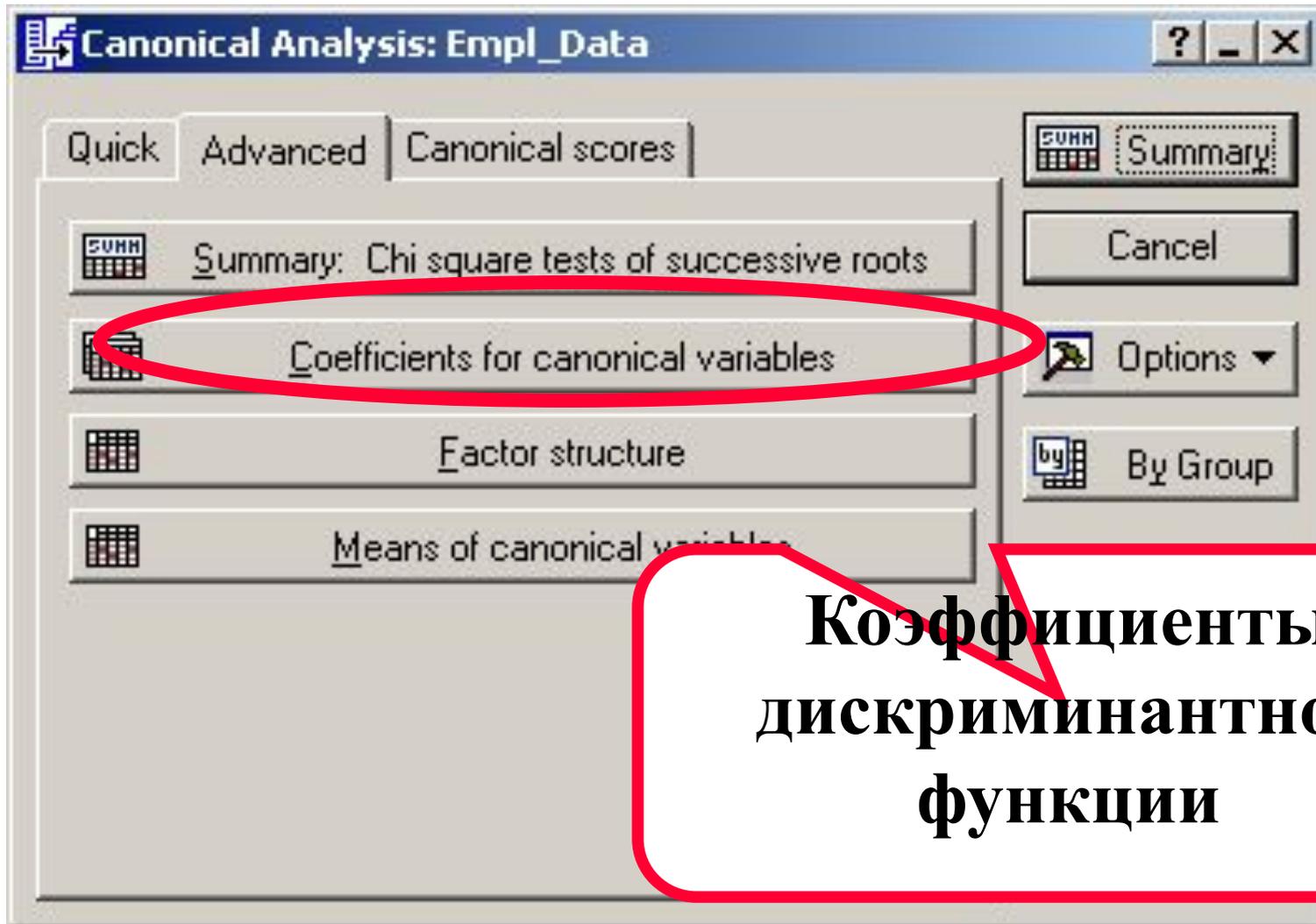
Canonical Analysis (Advanced):

Chi-Square Tests with Successive Roots Removed (empl_data.sta)

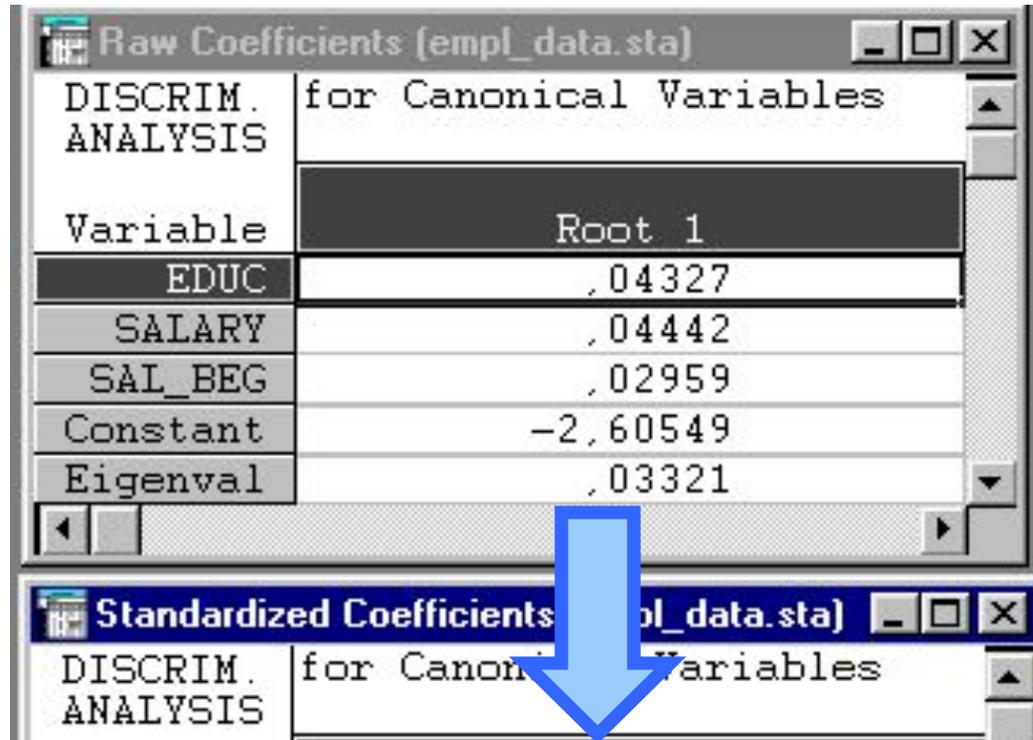
DISCRIM. ANALYSIS	Eigen-value	Canonical R	Wilks' Lambda	Chi-Sqr.	df	p-level
0	,033213	,179291	,967855	15,37274	3	,001527



Canonical Analysis (Advanced):



Canonical Analysis (Advanced):

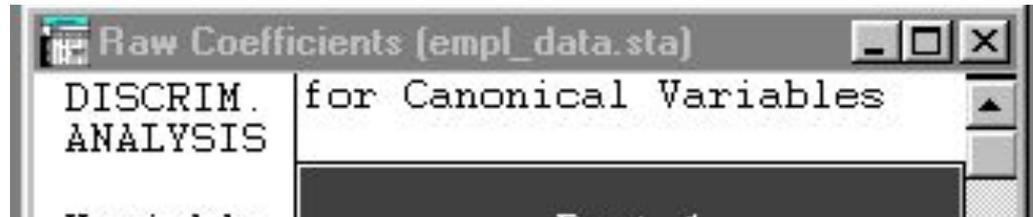


The screenshot shows the 'Raw Coefficients' window for a Canonical Analysis. The window title is 'Raw Coefficients (empl_data.sta)'. The main text reads 'DISCRIM. ANALYSIS for Canonical Variables'. Below this is a table with two columns: 'Variable' and 'Root 1'. The rows are: EDUC (.04327), SALARY (.04442), SAL_BEG (.02959), Constant (-2.60549), and Eigenval (.03321). A blue arrow points from the 'Raw Coefficients' window to the 'Standardized Coefficients' window below it.

Variable	Root 1
EDUC	.04327
SALARY	.04442
SAL_BEG	.02959
Constant	-2.60549
Eigenval	.03321

$$z = 0,043 * educ + 0,044 * salary + 0,030 * sal_beg - 2,605$$

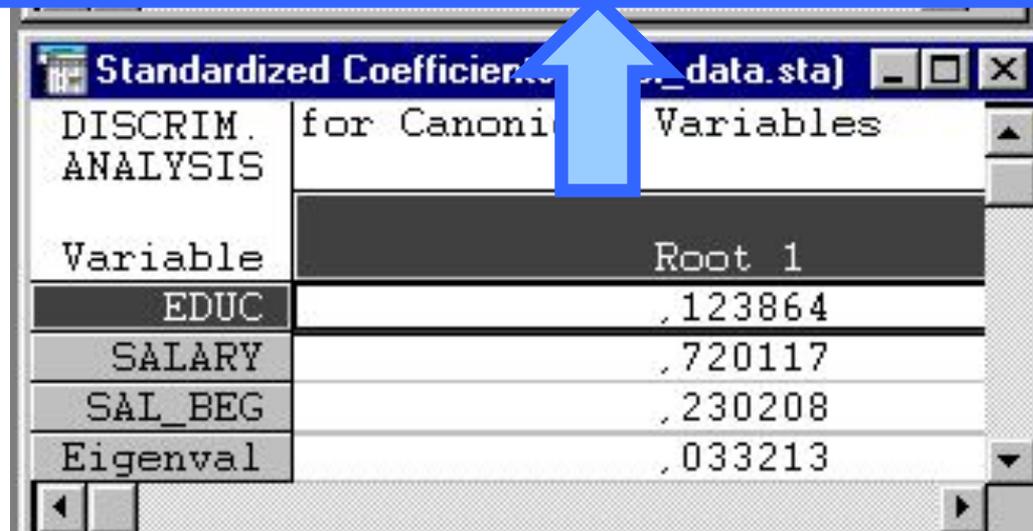
Canonical Analysis (Advanced):



Raw Coefficients (empl_data.sta)

DISCRIM. ANALYSIS	
for Canonical Variables	
Variable	Root 1

$$z = 0,124 * educ + 0,720 * salary + 0,230 * sal_beg$$

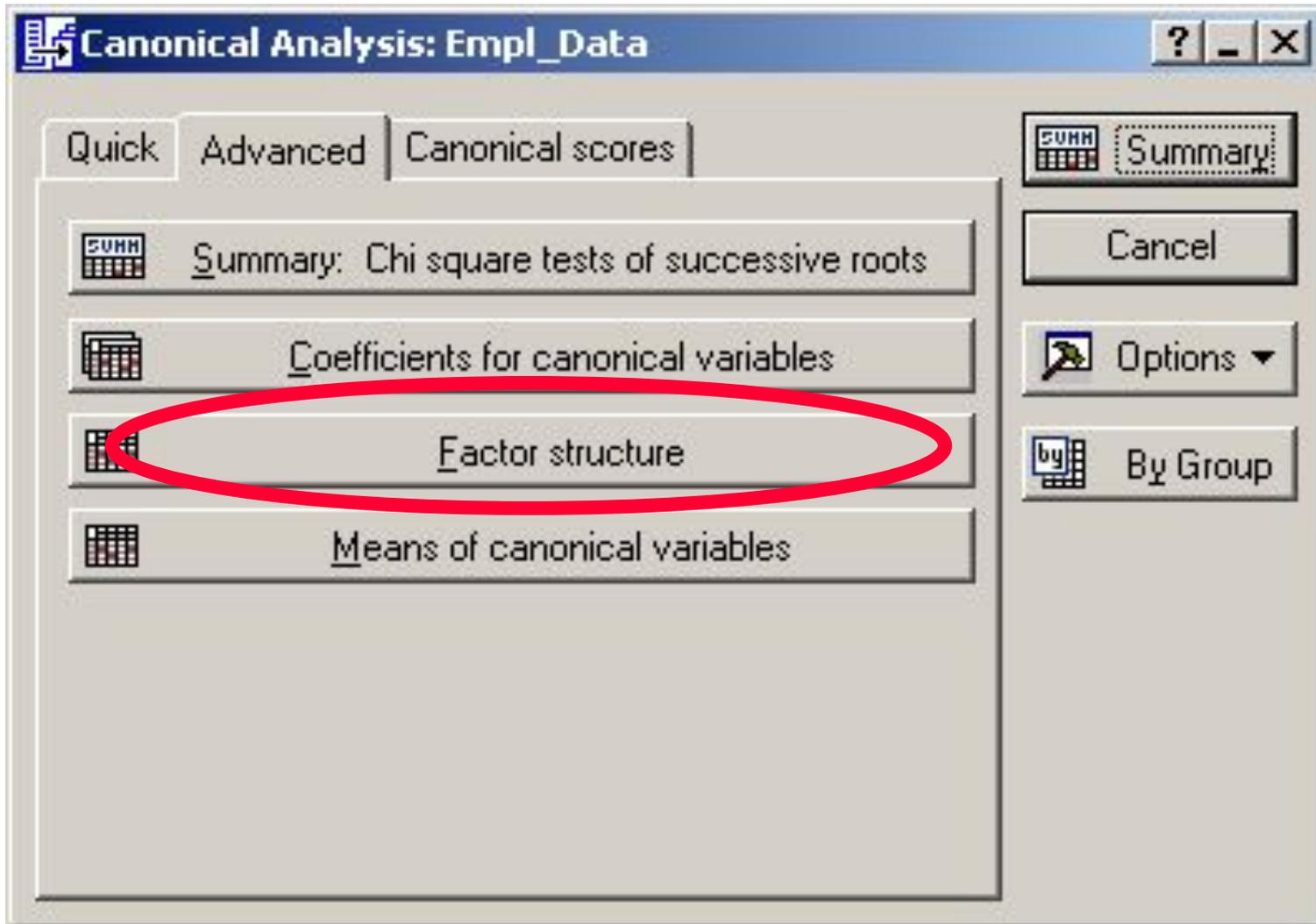


Standardized Coefficients (empl_data.sta)

DISCRIM. ANALYSIS	
for Canonical Variables	
Variable	Root 1
EDUC	.123864
SALARY	.720117
SAL_BEG	.230208
Eigenval	.033213

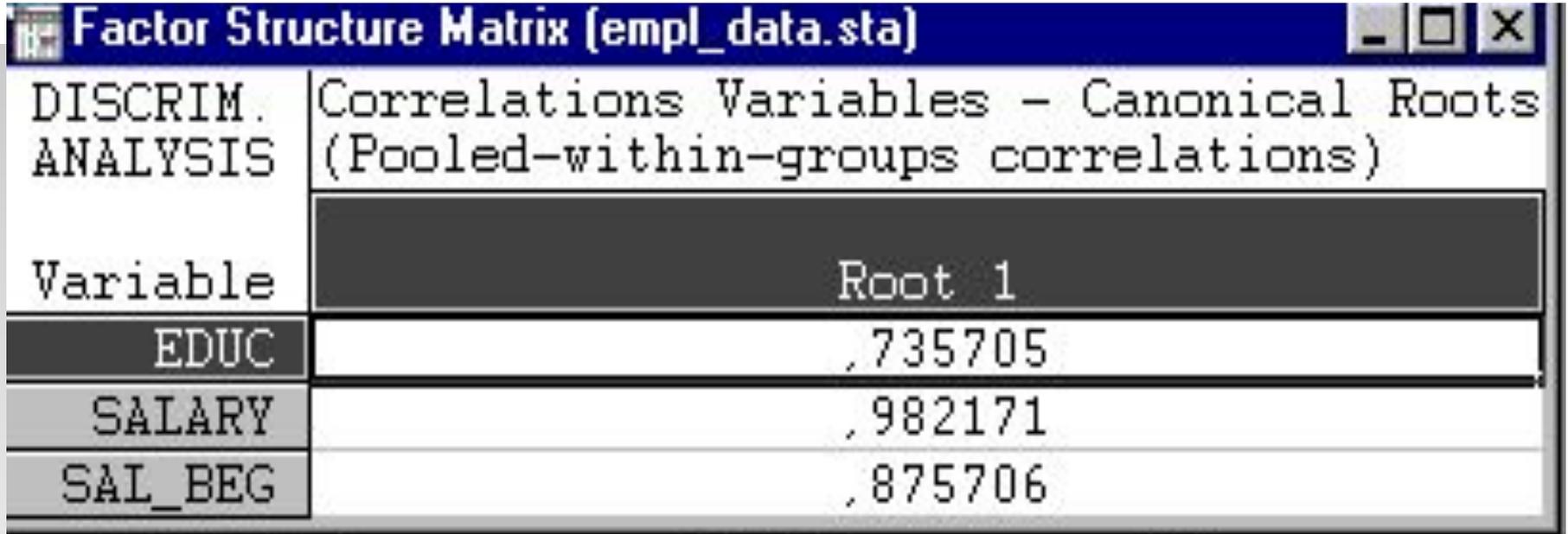


Canonical Analysis (Advanced):





Canonical Analysis (Advanced):



Factor Structure Matrix (empl_data.sta)

DISCRIM. ANALYSIS Correlations Variables - Canonical Roots
(Pooled-within-groups correlations)

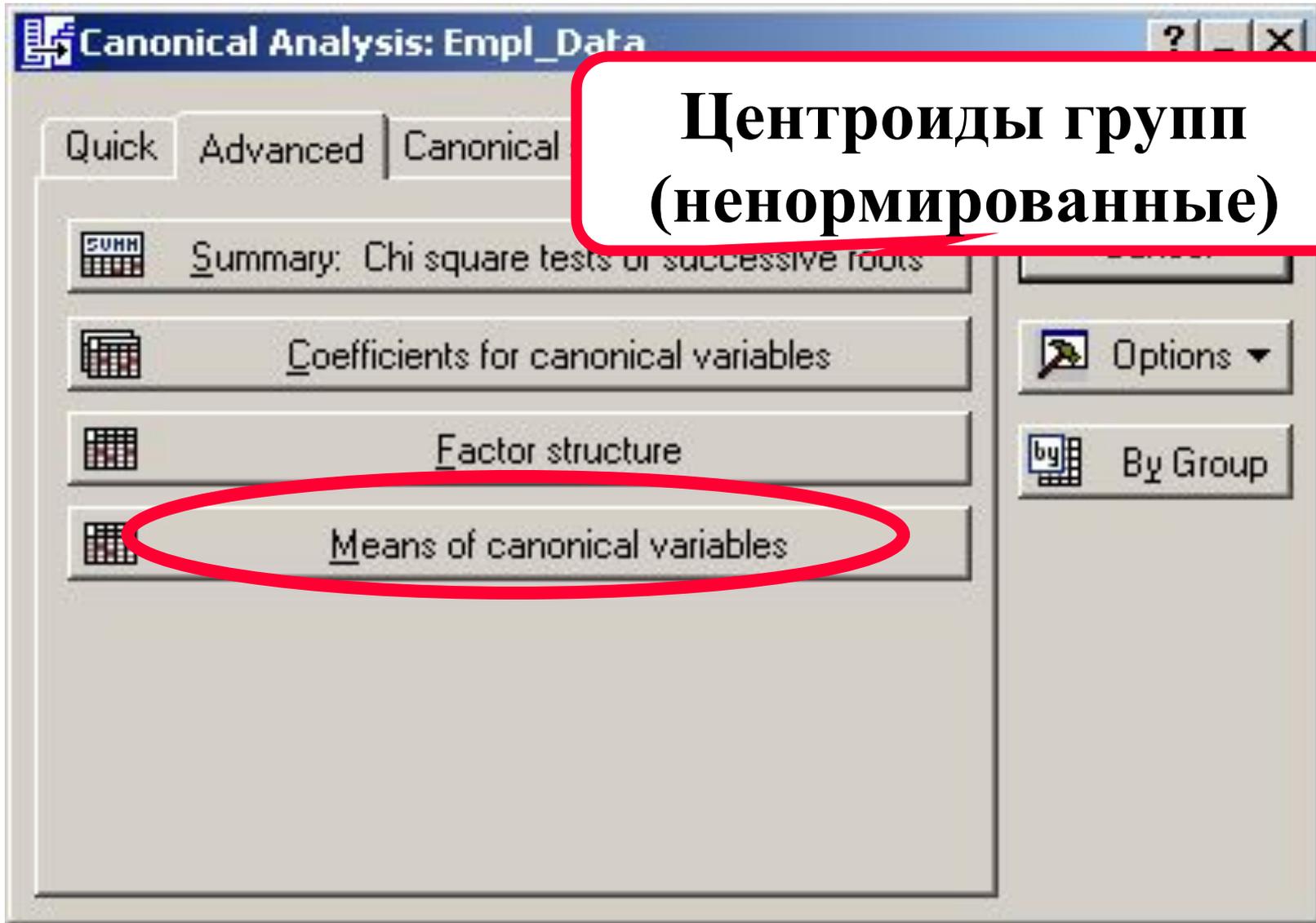
Variable	Root 1
EDUC	.735705
SALARY	.982171
SAL_BEG	.875706

**Корреляция переменных с
дискриминантной функцией**



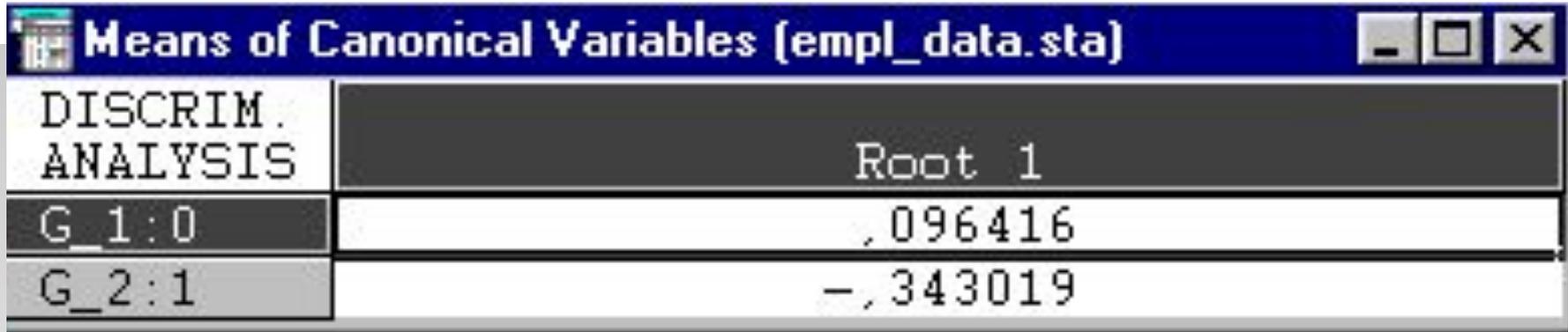
Canonical Analysis (Advanced):

**Центроиды групп
(ненормированные)**

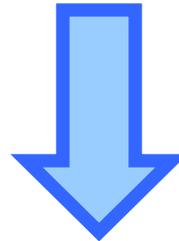




Canonical Analysis (Advanced):



Means of Canonical Variables (empl_data.sta)	
DISCRIM. ANALYSIS	Root 1
G_1:0	,096416
G_2:1	-,343019



$$z_{\Gamma P} = (0,096 - 0,342) / 2$$

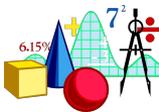
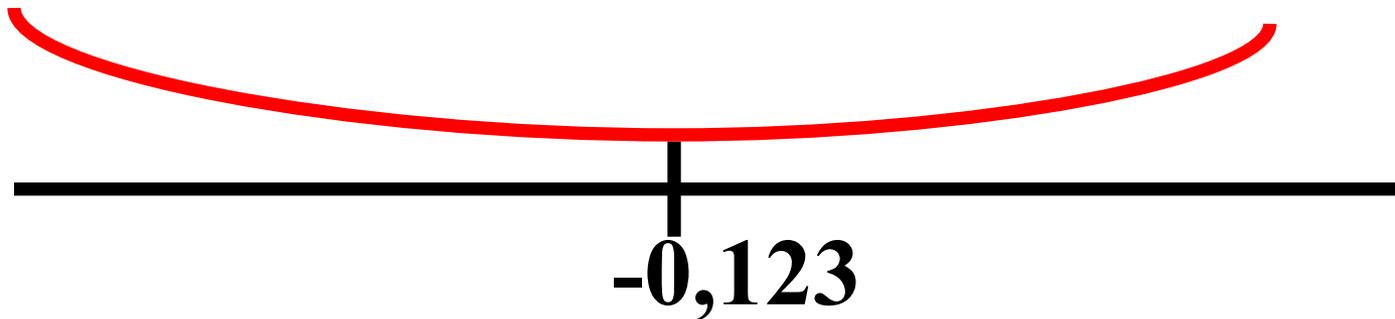


Пример для двух групп

$$z_{ГР} = (0,096 - 0,342) / 2 = -0,123$$

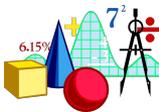
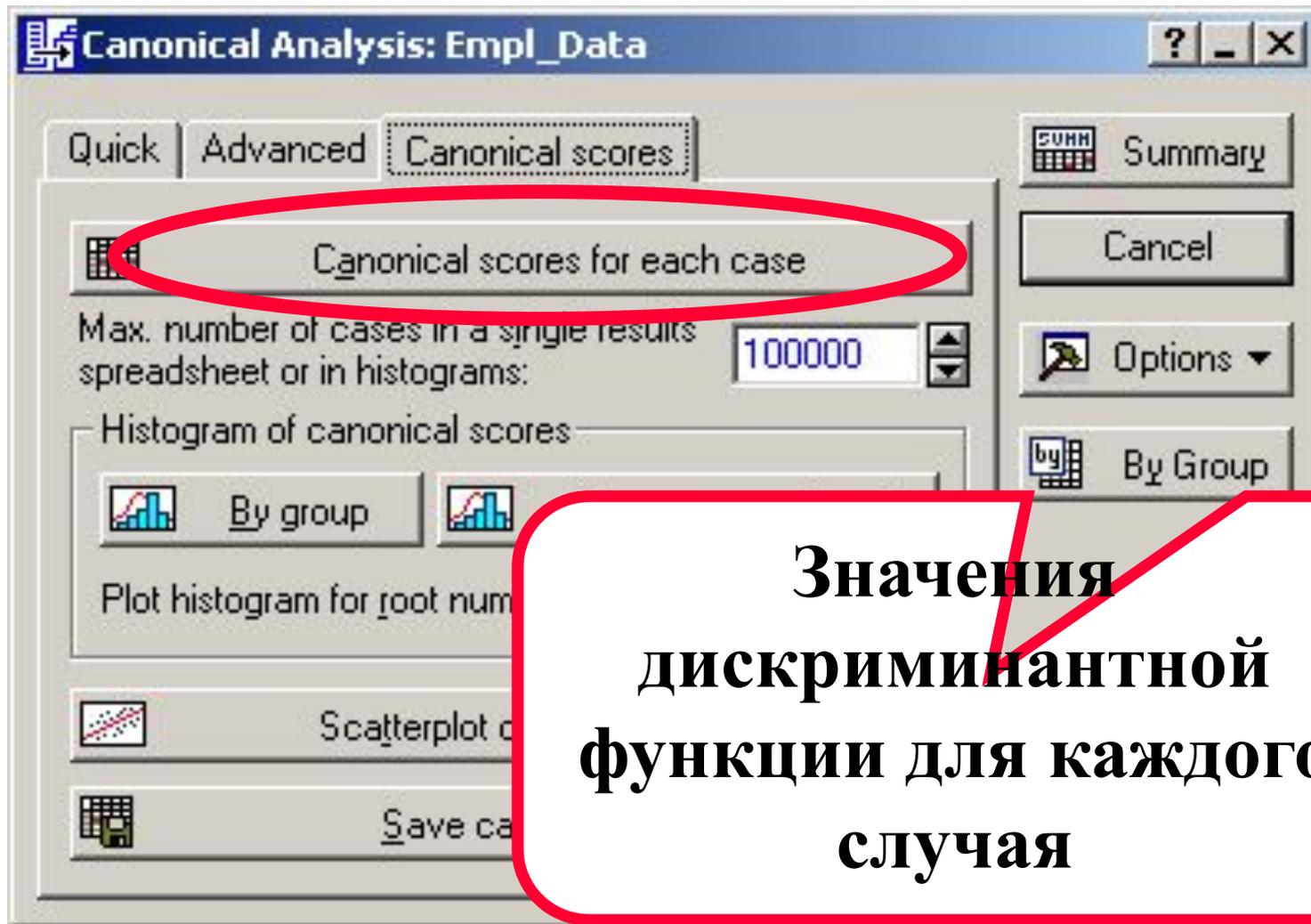
цветной

белый





Canonical Analysis (Canonical Scores):





Canonical Analysis (Canonical Scores):

STATISTICA: Discriminant Analysis - [Unstandardized Canonical Scores (empl_

File Edit View Analysis Graphs Options Window Help

G_1:0 Columns Rows

Continue...	Group	Root 1
1	G_1:0	1,37463
2	G_1:0	,42749
3	G_1:0	-,77826
4	G_1:0	-,89586
5	G_1:0	,66402
6	G_1:0	-,13095
7	G_1:0	,19764
8	G_1:0	-,82484
9	G_1:0	-,33972
10	G_1:0	-,62060
11	G_1:0	-,07887
12	G_2:1	-,64484
13	G_2:1	-,30200
14	G_2:1	,09996
15	G_1:0	-,47400
16	G_1:0	,17009
17	G_1:0	,50873
18	G_1:0	3,50979
19	G_1:0	,21454





Результаты анализа (Classification):

Discriminant Function Analysis Results

Number of variables in the model: 5
Wilks' Lambda: ,9681115 approx. eigenvalue = 5,187

Функции классификации

Quick | **Advanced** | Classification

Classification functions

Use selection conditions to classify selected cases only **SELECT CASES** 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

Summary

Cancel

Options

By Group





Результаты анализа (Classification):

Classification Functions; grouping: MINORITY (empl_data.sta)

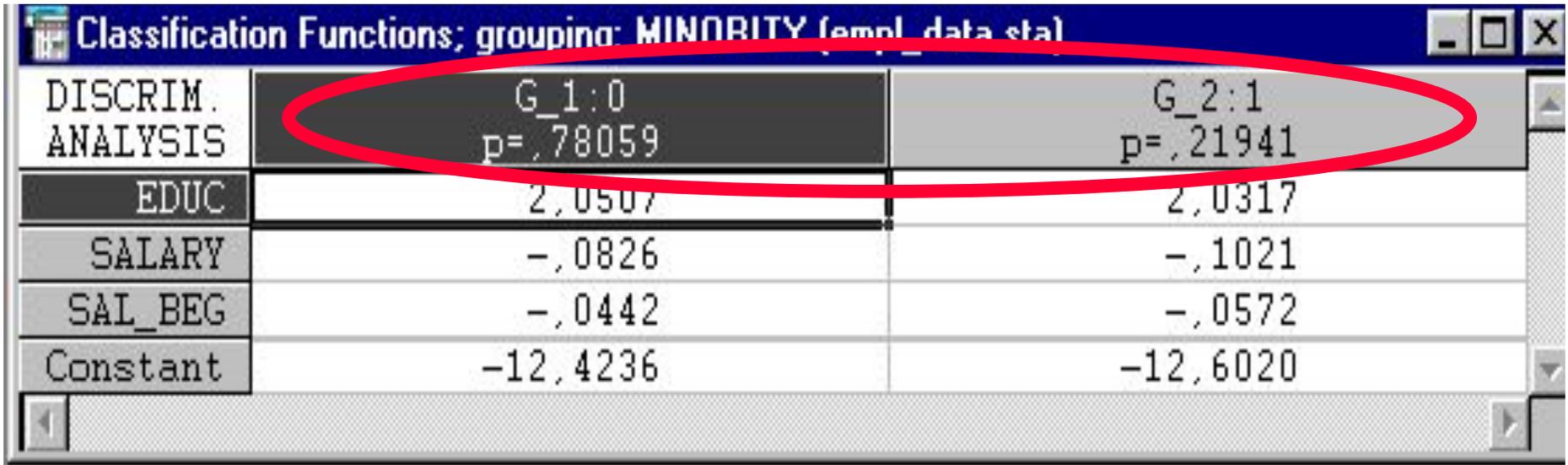
DISCRIM. ANALYSIS	G_1:0 p=,78059	G_2:1 p=,21941
EDUC	2,0507	2,0317
SALARY	-,0826	-,1021
SAL_BEG	-,0442	-,0572
Constant	-12,4236	-12,6020

Значения этих функций вычисляются для каждой группы и служат для прямой классификации. Случай попадает в группу, для которой у него получается наибольшее значение





Результаты анализа(Classification):



Classification Functions; grouping: MINORITY (empl_data.sta)

DISCRIM. ANALYSIS	G_1:0 p=,78059	G_2:1 p=,21941
EDUC	2,0507	2,0317
SALARY	-,0826	-,1021
SAL_BEG	-,0442	-,0572
Constant	-12,4236	-12,6020

Априорные вероятности попасть к данную группу (по умолчанию вычисляются исходя из размеров группы)





Результаты анализа:

Discriminant Function Analysis Results: Empl_Data

Number of variables in the model: 2

Wilks' Lambda:

Это очень полезная матрица!

Quick | Adva

Classification functions

Use selection conditions to classify selected cases only:

Classification matrix

Classification of cases

Squared Mahalanobis distances

Posterior probabilities

Save scores

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





Результаты анализа (Classification):

DISCRIM. ANALYSIS	Rows: Observed classifications Columns: Predicted classifications		
Group	Percent Correct	G_1:0 p=,78059	G_2:1 p=,21941
G_1:0	100,0000	370	0
G_2:1	0,0000	104	0
Total	78,0591	474	0

Очень важный показатель! Процент правильно предсказанных значений





Результаты анализа (Classification):

Discriminant Function Analysis Results: Empl_Data

Number of variables in the model: 3

Wilks' Lambda: ,9681115 approx. F (3,470) = 5,160421 p < ,0016

Quick | Advanced | **Classification**

Classification functions

Use selection conditions to classify selected cases only **SELECT CASES** 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

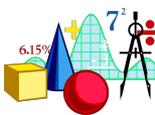
Max. number of cases in a single results spreadsheet: 100000

Summary

Cancel

Options

By Group





Результаты анализа (Classification):

STATISTICA: Discriminant Analysis - [Classification of Cases (empl_data.sta)]

File Edit View Analysis Graphs Options Window Help

G_1:0 Columns Rows

Incorrect classifications are marked with *

Case	Observed Classif.	1 p=,78059	2 p=,21941
1	G_1:0	G_1:0	G_2:1
2	G_1:0	G_1:0	G_2:1
3	G_1:0	G_1:0	G_2:1
4	G_1:0	G_1:0	G_2:1
5	G_1:0	G_1:0	G_2:1
6	G_1:0	G_1:0	G_2:1
7	G_1:0	G_1:0	G_2:1
8	G_1:0	G_1:0	G_2:1
9	G_1:0	G_1:0	G_2:1
10	G_1:0	G_1:0	G_2:1
11	G_1:0	G_1:0	G_2:1
* 12	G_2:1	G_1:0	G_2:1
* 13	G_2:1	G_1:0	G_2:1
* 14	G_2:1	G_1:0	G_2:1
15	G_1:0	G_1:0	G_2:1
16	G_1:0	G_1:0	G_2:1
17	G_1:0	G_1:0	G_2:1
18	G_1:0	G_1:0	G_2:1
19	G_1:0	G_1:0	G_2:1





Результаты анализа (Classification):

Discriminant Function Analysis Results: Empl_Data

Number of variables in the model: 3

Wilks' Lambda: ,9681115 approx. F (3,470) = 5,160421 p < ,0016

Quick | Advanced | **Classification**

Classification functions

Use selection conditions to classify selected cases only **SELECT CASES** 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

Summary

Cancel

Options

By Group





Пример для двух групп

STATISTICA: Discriminant Analysis - [Posterior Probabilities (empl_data.sta)]

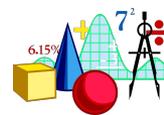
File Edit View Analysis Graphs Options Window Help

G_1:0 Columns Rows

DISCRIM. ANALYSIS

Incorrect classifications are marked with *

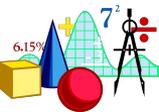
Case	Observed Classif.	G_1:0 p=,78059	G_2:1 p=,21941
1	G_1:0	,872957	,127043
2	G_1:0	,819232	,180768
3	G_1:0	,727366	,272634
4	G_1:0	,716999	,283001
5	G_1:0	,834117	,165883
6	G_1:0	,780015	,219985
7	G_1:0	,803789	,196211
8	G_1:0	,723288	,276712
9	G_1:0	,763869	,236131
10	G_1:0	,740887	,259113
11	G_1:0	,783916	,216084
* 12	G_2:1	,738837	,261163
* 13	G_2:1	,766846	,233154
* 14	G_2:1	,796931	,203069
15	G_1:0	,753060	,246940
16	G_1:0	,801873	,198127
17	G_1:0	,824458	,175542
18	G_1:0	,946120	,053880
19	G_1:0	,804957	,195043





Пример для трех групп

Посмотрим, можем ли мы предсказать, на какой должности работает человек по его зарплате, образованию и принадлежности к национальному меньшинству.

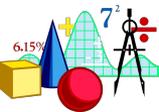




Пример для трех групп

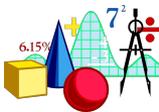
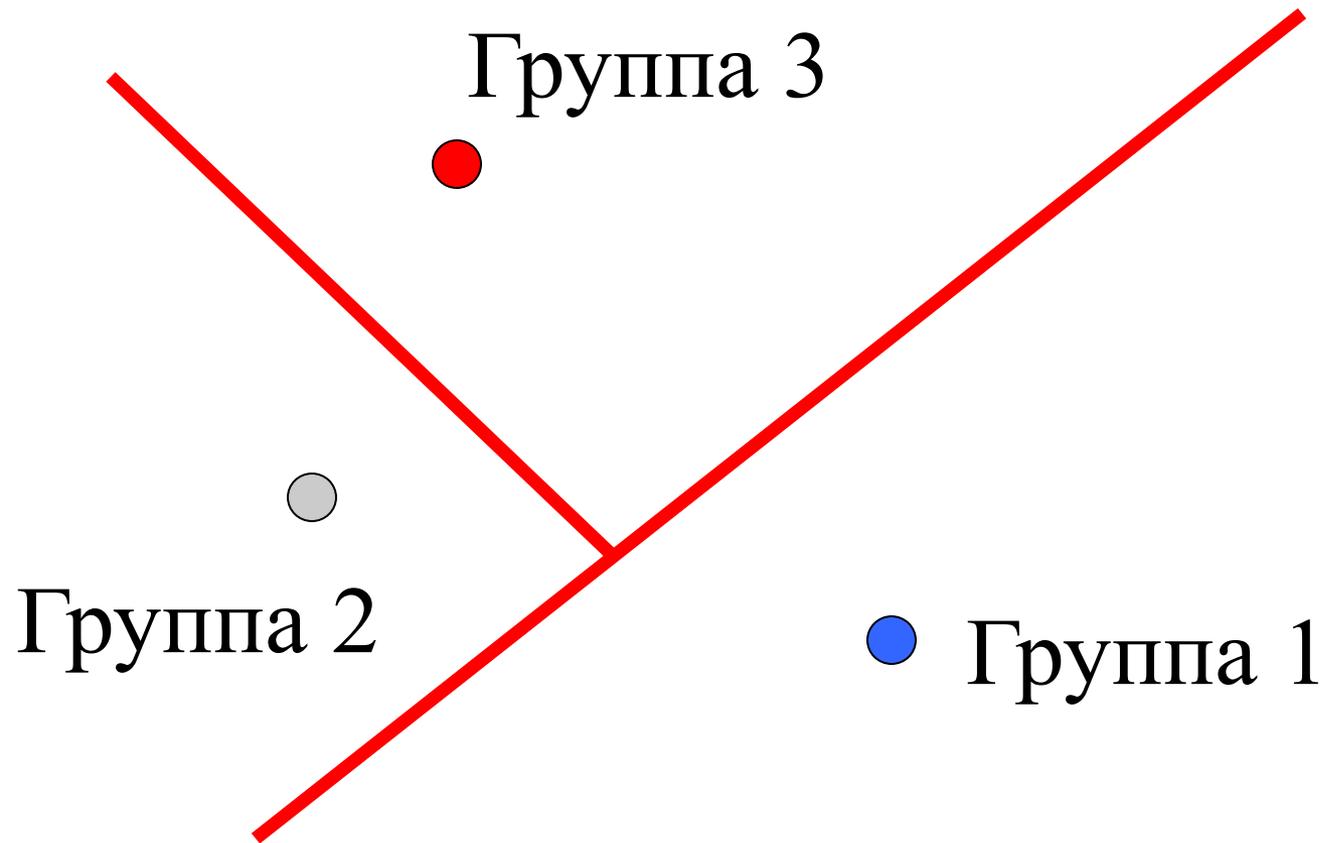
**В этом случае одной
дискриминантной функцией не
обойдешься!**

Их будет две.





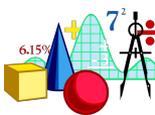
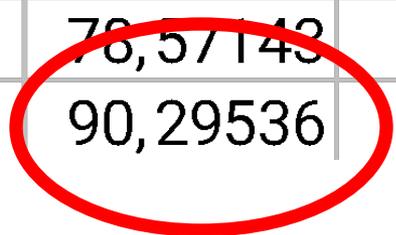
Пример для трех групп





Пример для трех групп

Classification Matrix (empl_data.sta)				
Rows: Observed classifications				
Columns: Predicted classifications				
	Percent	G_1:1	G_2:2	G_3:3
	Correct	p=,76582	p=,05696	p=,17722
G_1:1	97,79614	355	6	2
G_2:2	25,92593	20	7	0
G_3:3	78,57143	18	0	66
Total	90,29536	393	13	68





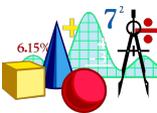
Пример для трех групп

Discriminant Function Analysis Summary (empl_data.sta)

No. of vars in model: 4; Grouping: JCAT (3 grps)

Wilks' Lambda: ,25680 approx. F (8,936)=113,88 p<0,0000

	Wilks'	Partial	F-remove			1-Toler.
	Lambda	Lambda	(2,468)	p-level	Toler.	(R-Sqr.)
EDUC	0,290012	0,885496	30,25871	4,38E-13	0,805286	0,194714
SALARY	0,324602	0,791136	61,77707	1,55E-24	0,731865	0,268135
SAL_BEG	0,299848	0,85645	39,22078	1,79E-16	0,765182	0,234818
MINORITY	0,264071	0,972483	6,62129	0,00146	0,996228	0,003772





Пример для трех групп

Chi-Square Tests with Successive Roots Removed (empl_data.sta)

	Eigen- value	Canonica R	Wilks' Lambda	Chi-Sqr.	df	p-level
0	2,388448	0,839571	0,256805	638,2568	8	0
1	0,149202	0,36032	0,870169	65,29217	3	4,49E-14



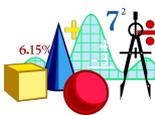


Пример для трех групп

Factor Structure Matrix (empl_data.sta)

DISCRIM. ANALYSIS Correlations Variables - Canonical Roots (Pooled-within-groups correlations)

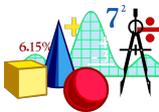
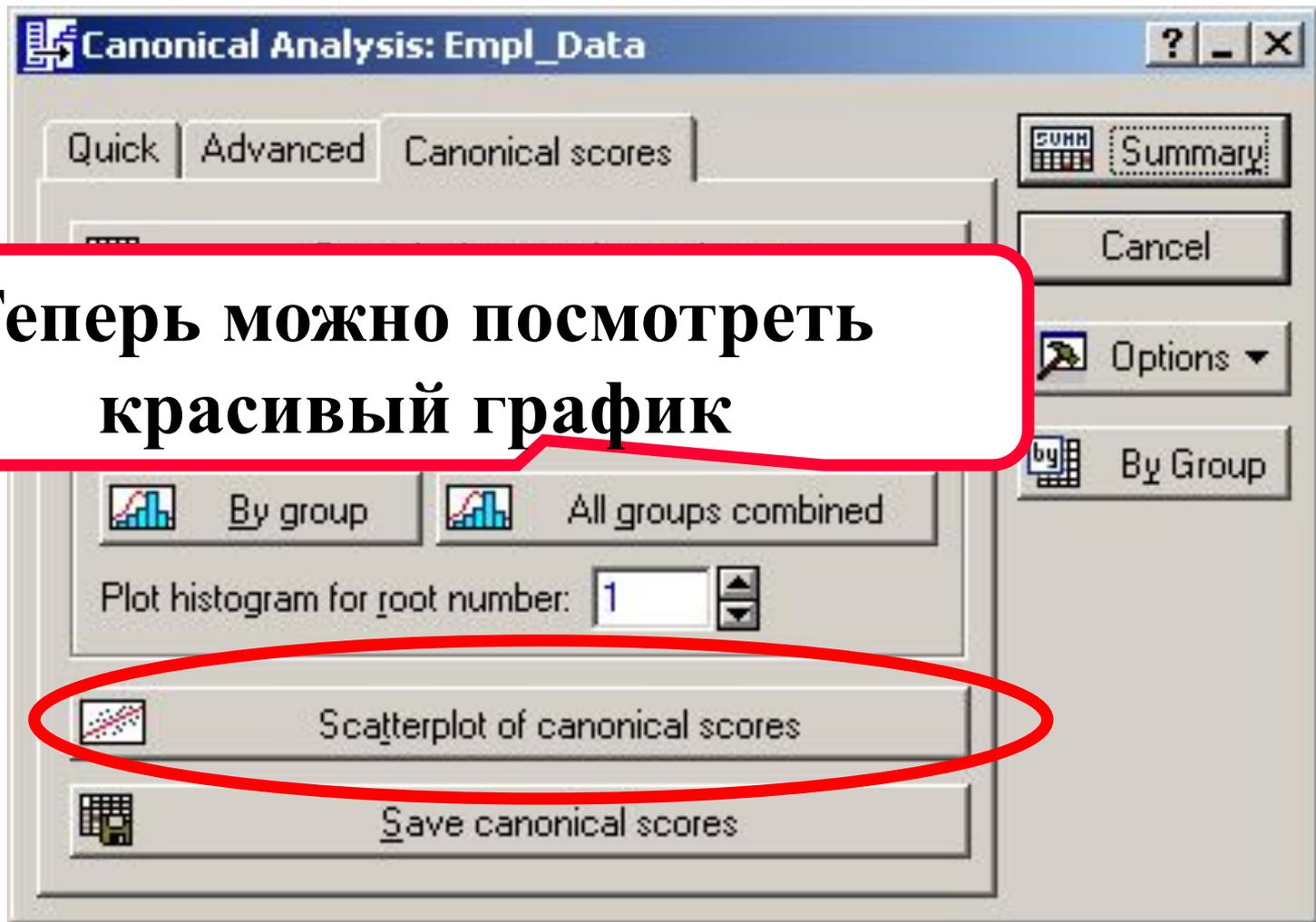
Variable	Root 1	Root 2
EDUC	.542780	.736306
SALARY	.932154	-.189787
SAL_BEG	.862088	-.124981
MINORITY	-.136299	-.365001





Пример для трех групп

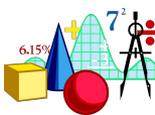
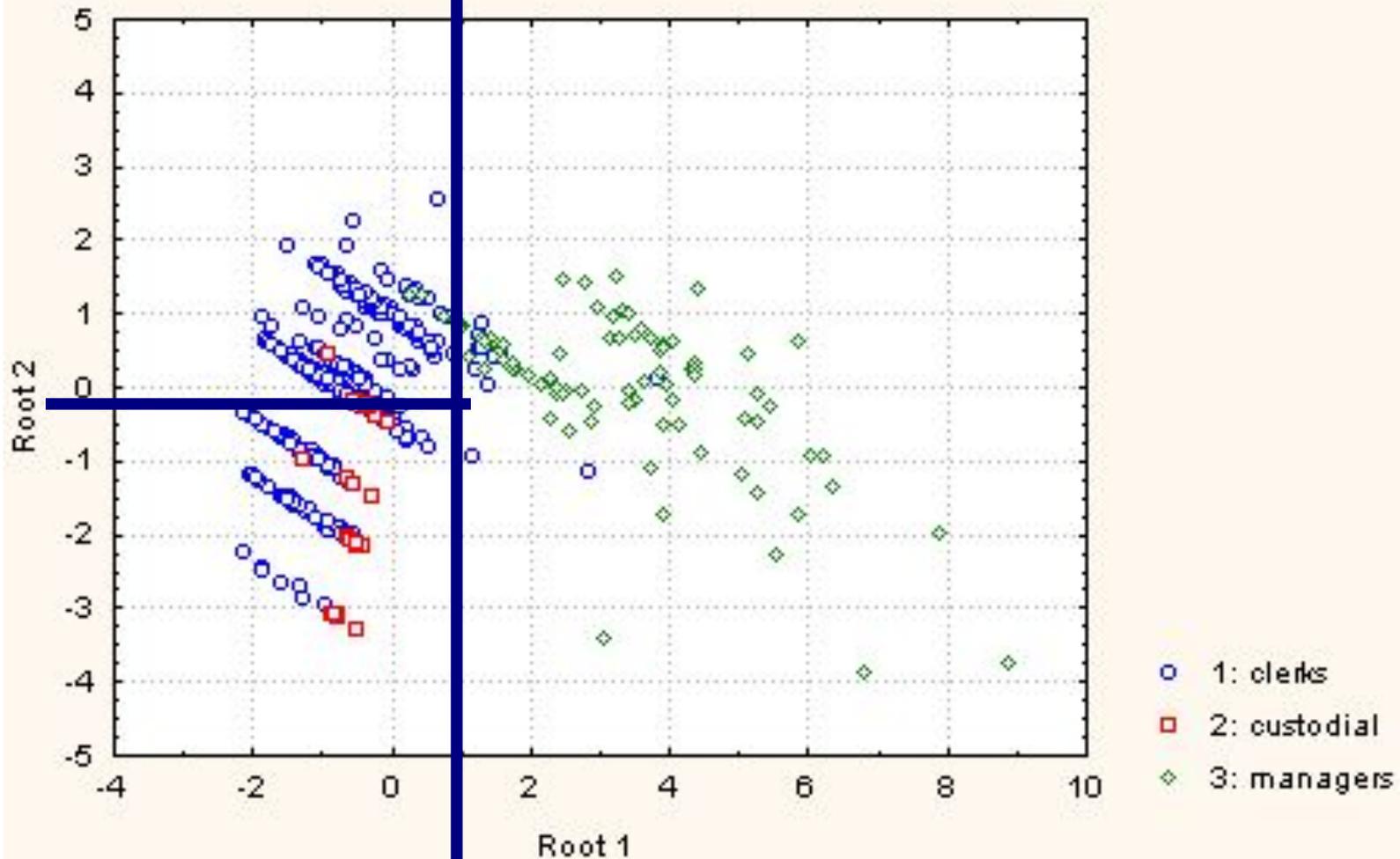
Теперь можно посмотреть
красивый график





Пример для трех групп

Root 1 vs. Root 2

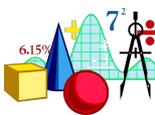
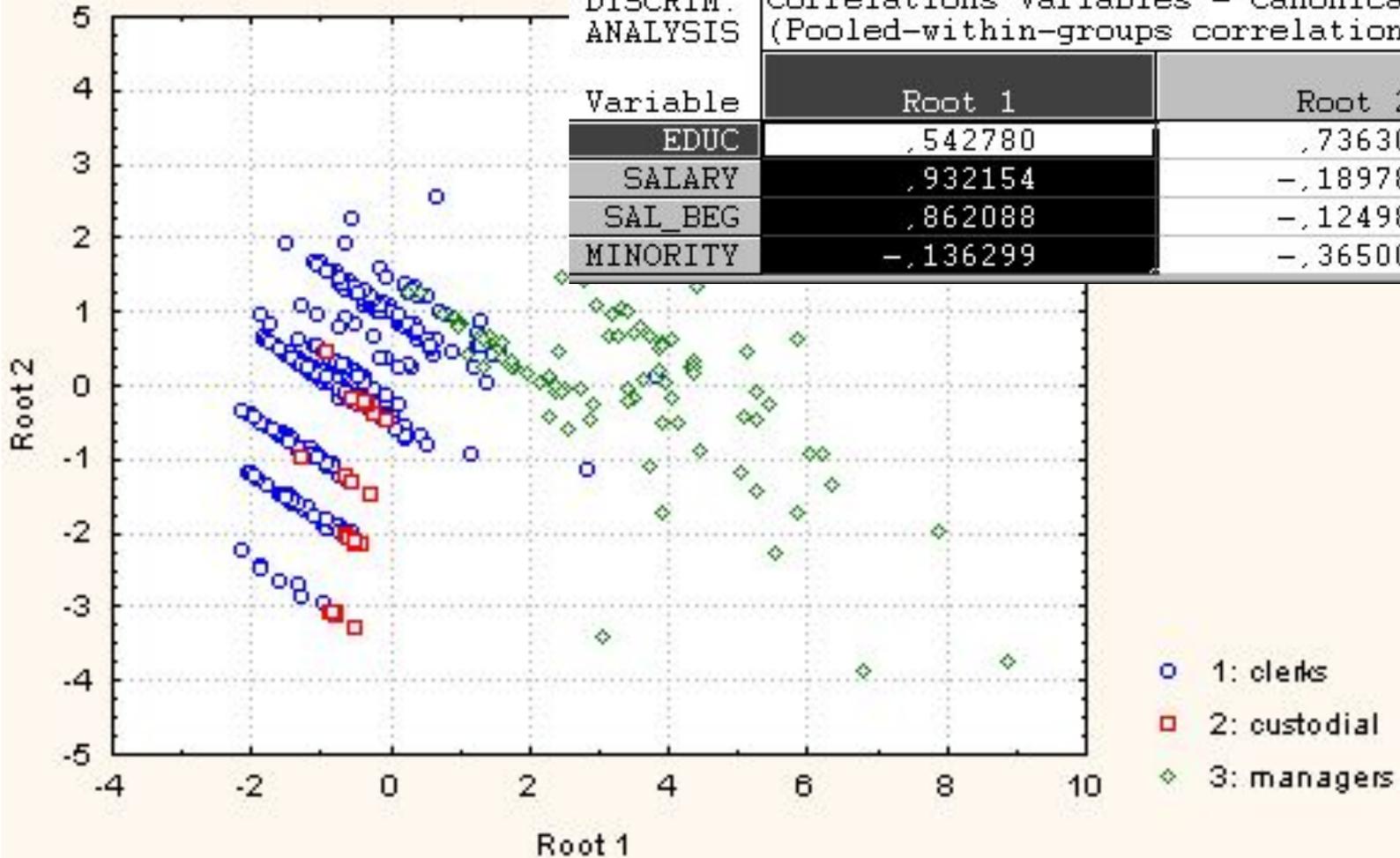




Пример для трех групп

Root 1 Factor Structure Matrix (empl_data.sta)

Variable	Root 1	Root 2
EDUC	,542780	,736306
SALARY	,932154	-,189787
SAL_BEG	,862088	-,124981
MINORITY	-,136299	-,365001

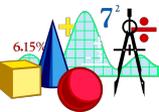




Результаты анализа

Мы можем

- 1) оценить, насколько НП определяют ЗИИ (т.е оценить нашу модель)
- 2) делать предсказания (по значениям НП определять, в какую группу попадет объект или индивид)





Как делать прогноз?

STATISTICA: Discriminant Analysis - [Data: Empl_Data.STA 9v * 475c]

File Edit View Analysis Graphs Options Window Help

-9999, Vars Cases

TEXT VALU	1 ID	2 GENDER	3 EDUC	4 JCAT	5 SALARY	6 SAL_BEG	7 JTIME	8 PREVEX	9 MINORITY
450	450,000	m	19	3	55,000	34,980	65	129	0
451	451,000	m	15	1	28,500	14,250	65	20	0
452	452,000	m	12	1	28,800	18,000	65	210	0
453	453,000	m	15	1	24,450	15,750	65	338	0
454	454,000	m	19	3	90,625	31,250	65	18	0
455	455,000	m	16	3	43,650	19,500	65	19	0
456	456,000	m	19	3	75,000	42,510	65	54	0
457	457,000	m	15	1	31,650	14,250	65	10	0
458	458,000	m	19	3	61,875	28,740	65	26	0
459	459,000	f	12	1	21,750	11,250	65	0	0
460	460,000	f	12	1	22,500	12,750	65	24	0
461	461,000	f	8	1	21,600	13,500	65	173	0
462	462,000	f	16	3	34,410	19,500	65	79	0
463	463,000	f	15	1	20,700	14,250	65	241	0
464	464,000	m	19	3	47,550	33,000	64	27	0
465	465,000	m	12	1	33,900	16,500	64	106	0
466	466,000	f	12	1	23,400	13,500	64	198	0
467	467,000	f	16	1	32,850	19,500	64	20	0
468	468,000	f	16	3	55,750	19,980	64	36	0
469	469,000	f	15	1	25,200	13,950	64	57	0
470	470,000	m	12	1	26,250	15,750	64	69	1
471	471,000	m	15	1	26,400	15,750	64	32	1
472	472,000	m	15	1	39,150	15,750	63	46	0
473	473,000	f	12	1	21,450	12,750	63	139	0
474	474,000	f	12	1	29,400	14,250	63	9	0
475			18		7,000	1,000			0

Ready Output:OFF Set:OFF Weight:OFF





Как делать прогноз?

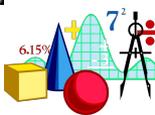
Discriminant Function Analysis Results

Number of variables in the model: 3

Wilks' Lambda: ,9678548 approx. F (3, 470) = 5,203344 p < ,0015

Variables <u>i</u> n the model	Classification matrix	
Variables <u>n</u> ot in the model	Classification of cases	
Distances between groups	Squared Mahalanobis distances	
Summary of stepwise analysis	Posterior probabilities	
Canonical analysis & graphs	Save classif/distances/probs	
Classification <u>f</u> unctions	A priori classification probabilities <input checked="" type="radio"/> Proportional to group sizes <input type="radio"/> Same for all groups <input type="radio"/> User defined	

NOTE: Use case selection conditions to classify selected cases only





Как делать прогноз?

DISCRIM. ANALYSIS Incorrect classifications are marked with *

Case	Observed Classif.	1 p=,76582	2 p=,05696	3 p=,17722
469	G_1:1	G_1:1	G_2:2	G_3:3
470	G_1:1	G_1:1	G_2:2	G_3:3
471	G_1:1	G_1:1	G_2:2	G_3:3
472	G_1:1	G_1:1	G_2:2	G_3:3
473	G_1:1	G_1:1	G_2:2	G_3:3
474	G_1:1	G_1:1	G_2:2	G_3:3
475	---	G_1:1	G_2:2	G_3:3

Posterior Probabilities (empl_data.sta)

DISCRIM. ANALYSIS Incorrect classifications are marked with *

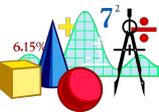
Case	Observed Classif.	G_1:1 p=,76582	G_2:2 p=,05696	G_3:3 p=,17722
465	G_1:1	,955130	,043518	,001352
466	G_1:1	,983686	,016263	,000051
467	G_1:1	,990207	,002422	,007371
468	G_3:3	,398281	,006974	,594745
469	G_1:1	,997685	,002134	,000181
470	G_1:1	,891187	,108737	,000076
471	G_1:1	,986535	,013289	,000176
472	G_1:1	,985285	,007423	,007292
473	G_1:1	,986566	,013408	,000026
474	G_1:1	,972155	,027600	,000245
475	---	,999970	,000029	,000001





Пример (реальный)

Проект: Можно ли предсказать тип преступника (насильственный, корыстный или корыстно-насильственный) по результатам тестов Кеттела и Леонгарда-Шмишека?

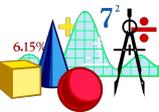




Пример

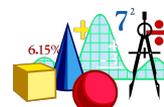
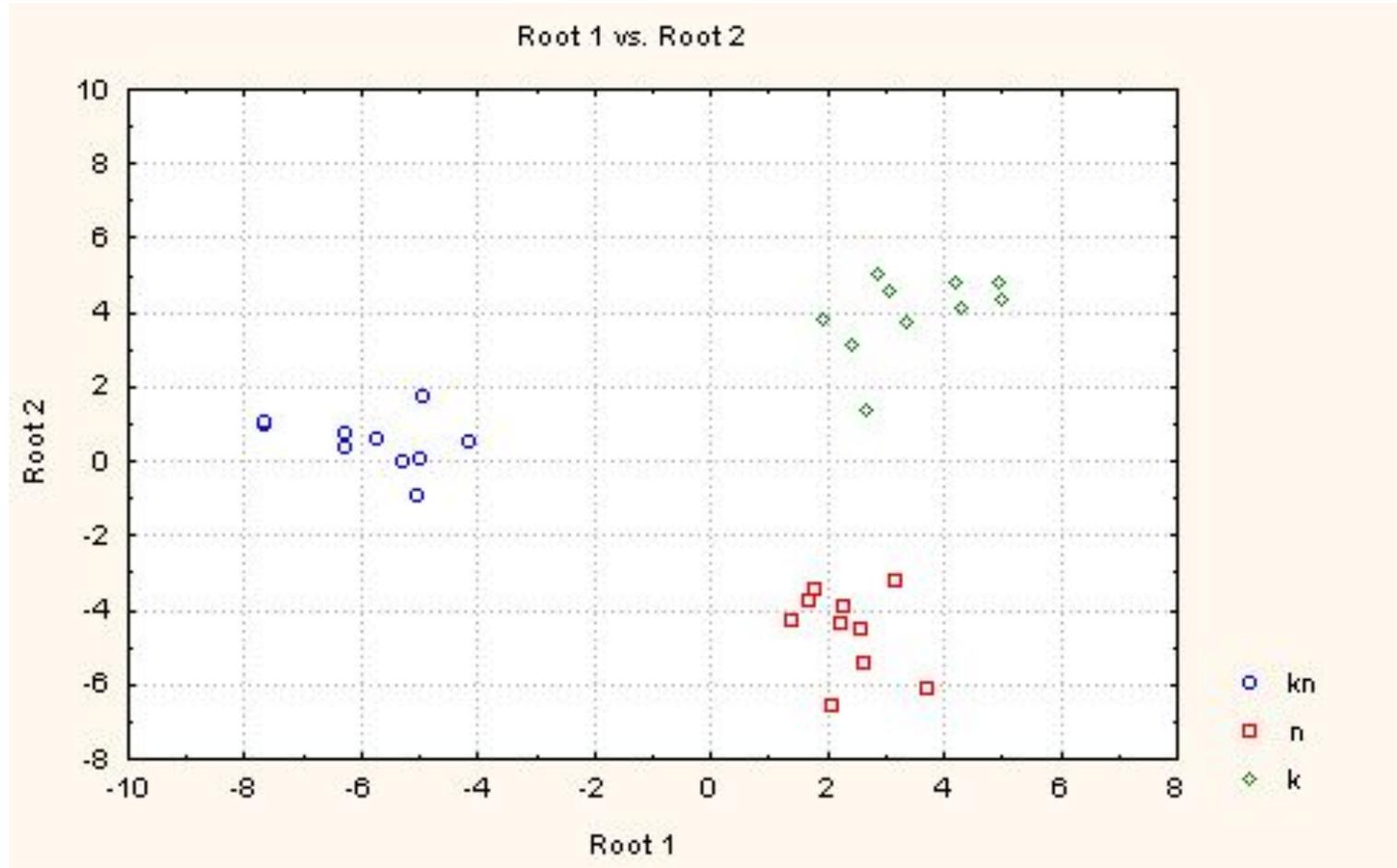
- 1) дискриминантный анализ по всем переменным.
- 2) прямой пошаговый дискриминантный анализ.

Получились совершенно потрясающие результаты:





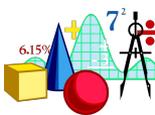
Пример (результаты)





Пример (результаты)

Chi-Square Tests with Successive Roots Removed (issled.sta)						
	Eigen- value	Canonici R	Wilks' Lambda	Chi-Sqr.	df	p-level
0	19,02697	0,974714	0,003411	96,57309	42	0
1	13,63901	0,965241	0,068311	45,62273	20	0,00091





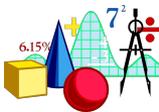
Пример (результаты –ГМ!)

Factor Structure Matrix (issled.st
Correlations Variables - Canonical
(Pooled-within-groups correlations)

	Root 1	Root 2
F3	0,11	0,19
A	0,08	0,21
F4	-0,11	0,07
H	0,08	0,13
EKZ	-0,02	-0,01
F	0	0,02
Q3	0,04	0
Q1	0,01	0,11
TREV	0,03	-0,03
DEM	0,06	0,08
VOZB	0,03	-0,08
O	-0,03	0,07
PED	0,03	0,01
MD	0,05	-0,08
ZASTR	0	0
N	0,05	0,11
B	0,03	-0,01
M	-0,11	0,01
L	0,04	-0,03
EMOT	-0,04	-0,02
CIKL	0	0,01

Standardized Coefficients (issled
for Canonical Variables

	Root 1	Root 2
F3	0,47	0,27
A	-1	2,96
F4	-1,34	2,94
H	3,88	0,31
EKZ	1,04	-1,15
F	0,12	-1,02
Q3	-1,33	1,64
Q1	0,75	-1,49
TREV	1,94	0,29
DEM	-0,63	1,79
VOZB	-0,46	-1,06
O	-3,32	1,82
PED	-5,26	-0,11
MD	6,32	-0,85
ZASTR	3,09	-0,6
N	4,6	-1,4
B	0,69	-0,49
M	-3,13	-0,14
L	-1,7	0,79
EMOT	2,29	-0,64
CIKL	-1,16	0,54





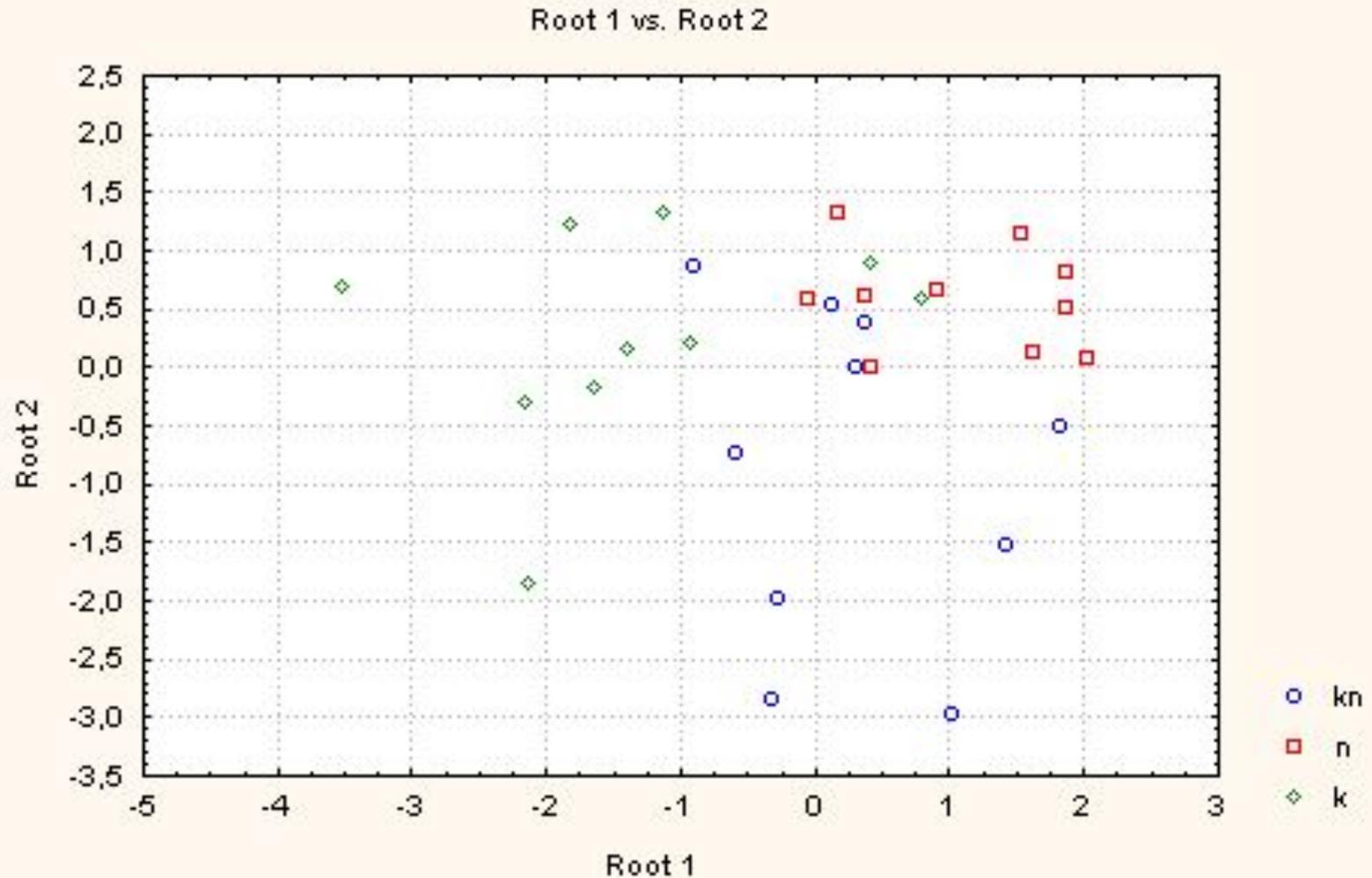
Пример (результаты –ГМ!)

Analysis of Variance (issled.sta)					
Marked effects are significant at $p < ,05000$					
	df	df			
	Effect	Error	F		p
MD	2	27	1,63		0,2139
A	2	27	9,32		0,0008
B	2	27	0,19		0,8242
C	2	27	2,15		0,1364
E	2	27	2,06		0,1472
F	2	27	0,07		0,9309
G	2	27	0,04		0,9574
H	2	27	4,82		0,0163
I	2	27	0,16		0,852
L	2	27	0,65		0,5321
M	2	27	2,85		0,0753
N	2	27	2,66		0,0886
O	2	27	1,2		0,3174
Q1	2	27	2,26		0,1237
Q2	2	27	0,42		0,6634
Q3	2	27	0,4		0,6747
Q4	2	27	4,82		0,0162
F1	2	27	5,83		0,0078
F2	2	27	1,56		0,2278
F3	2	27	10,2		0,0005
F4	2	27	4,18		0,0262
DEM	2	27	1,99		0,1559
ZASTR	2	27	0		1
PED	2	27	0,19		0,8253
VOZB	2	27	1,45		0,2531
GIPERT	2	27	3		0,0664
DISTIMN	2	27	1,61		0,2192
TREV	2	27	0,33		0,723
EKZ	2	27	0,18		0,8392
EMOT	2	27	0,52		0,5993
CIKL	2	27	0,02		0,9825





Пример (результаты –ГМ!)





Пример (результаты –ГМ!)

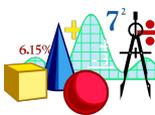
Classification Matrix (issled.sta)				
Rows: Observed classifications				
Columns: Predicted classifications				
	Percent	kn	n	k
	Correct	p=,33333	p=,33333	p=,33333
kn	60	6	3	1
n	90	1	9	0
k	80	0	2	8
Total	76,66666	7	14	9





Пример

Classification Matrix (issled.sta)				
Rows: Observed classifications				
Columns: Predicted classifications				
	Percent	kn	n	k
	Correct	p=,33333	p=,33333	p=,33333
kn	80	8	1	1
n	90	1	9	0
k	90	0	1	9
Total	86,66666	9	11	10

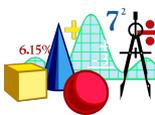




Пример

Factor Structure Matrix (issled.st
Correlations Variables - Canonical
(Pooled-within-groups correlations)

	Root 1	Root 2
A	-0,51	-0,17
H	-0,37	0,02
N	-0,27	-0,05
Q4	0,08	-0,8
F1	-0,41	-0,09
F3	-0,54	0,05
F4	0,03	-0,76
GIPERT	-0,29	0,08





Полезная литература

ПРОГРАММА STATISTICA

Боровиков В. Программа *STATISTICA* для студентов и инженеров. - Компьютер Пресс: Москва, 2001.

Электронный учебник по программе (StatSoft)

ПРОГРАММА SPSS

Наследов А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. – СПб. – Речь. – 2004.

Бююль А., Цефель П. SPSS: Искусство обработки информации. – СПб, «ЛианСофтЮп». –2001.





**К практическому занятию по регрессионному анализу
надо прочитать:**

- **Нестеренко А.И. и др. Прогноз тревожности у студенток на основании их типологических различий// ПЖ, 2003, т.24, № 6, с. 37-46**





Дискриминантный анализ

СПАСИБО

ЗА ВНИМАНИЕ!

