

ТЕМА 10

# ДИСКРИМИНАНТНЫЙ АНАЛИЗ

Лабораторные работы 19 и 20

# Лабораторная работа №19

## **Дискриминантный анализ: метод принудительного включения**

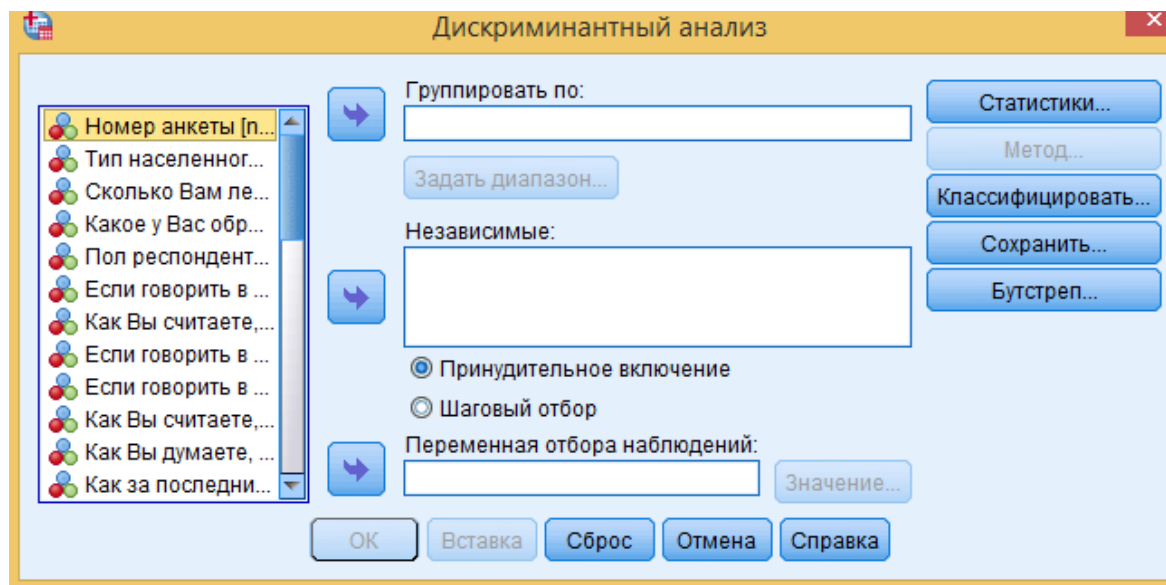
## Лабораторная работа №19

---

1. Откройте файл **Elections.sav**.
2. Спрогнозируйте, за кого из кандидатов проголосует большинство респондентов (переменная q13). При этом учтите, что многие пока не решили, за кого будут голосовать, затруднялись/уклонялись от ответа на вопрос.
3. Интерпретируйте полученные результаты.

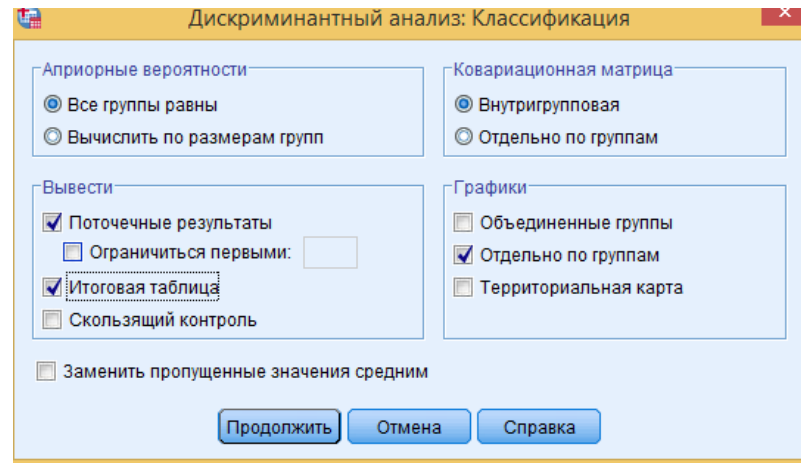
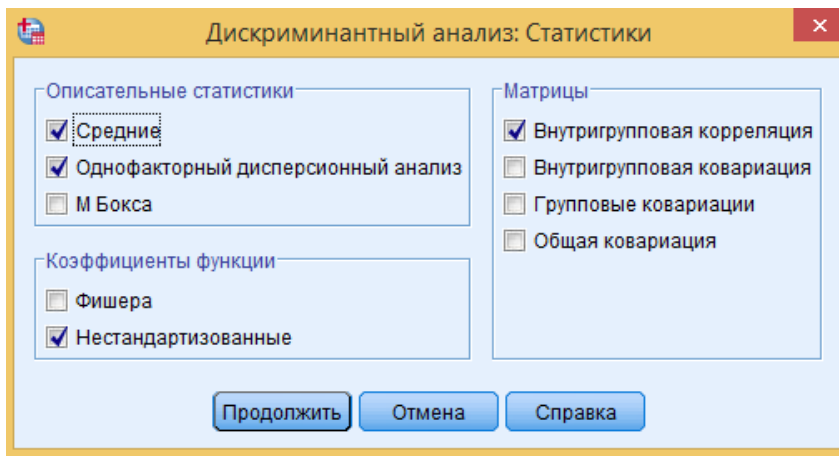
## Лабораторная работа №19

1. Команды «Анализ» → «Классификация» → «Дискриминантный анализ».
2. Зависимая переменная q13 («Если бы выборы...?») переносится в поле «Группировать по».
3. Задайте «Диапазон» от 1 до 3.
4. В качестве независимых переменных (факторных признаков) возьмем s1 (Возраст), s2 (Образование), q6\_2 (Как за последний год изменилась ваша жизнь? – Личный доход) и q8\_1 (По вашему мнению, как работают следующие органы власти: хорошо или плохо? – Президент) и перенесем в поле «Независимые переменные».
5. Оставьте метод «Принудительное включение».



## Лабораторная работа №19

6. Во вкладке «**Статистики**» отмечаем:
- в разделе «**Описательные статистики**»: «**Средние**», «**Однофакторный дисперсионный анализ**».
  - в разделе «**Коэффициенты функции**»: «**Нестандартизированные**».
  - в разделе «**Матрицы**»: «**Внутригрупповая корреляция**».
7. Во вкладке «**Классифицировать**» отмечаем:
- «**Все группы равны**»;
  - «**Внутригрупповая**»;
  - вывод «**Поточечных результатов**» и «**Итоговой таблицы**»;
  - «**Отдельно по группам**».



## Лабораторная работа №19

Анализ сводки обработки наблюдений

Невзвешенные наблюдения		N	Проценты
Допустимо		512	45,2
Исключено	Отсутствующие или выходящие за пределы диапазона коды групп	620	54,8
	По крайней мере одна дискриминирующая переменная	0	,0
	И отсутствующие или выходящие за пределы диапазоны коды групп, и по крайней мере одна дискриминирующая переменная	0	,0
	Всего	620	54,8
Всего		1132	100,0

- Всего 1132 наблюдения.
- 512 респондентов отметили, за кого будут голосовать на выборах, для этих наблюдений проводится дискриминантный анализ.
- 510 респондентов из 620 не отнесены ни к одной из трех групп, так как они затруднились ответить. Дискриминантный анализ позволит спрогнозировать их выбор на голосовании.

## Лабораторная работа №19

Статистика группы

Если бы выборы состоялись в следующее воскресенье, за кого из представленных кандидатов Вы бы, скорее всего, проголосовали?	Среднее значение	Стандартная отклонения	N валидных (по списку)	
			Невзвешенных	Взвешенных
Иванов И.И. Сколько Вам лет (укажите полное число лет)? Какое у Вас образование? Как за последний год изменилась Ваша жизнь и ее отдельные стороны? - Ваш личный доход По Вашему мнению, как работают следующие органы власти: хорошо или плохо? - Президент	45,85	16,207	201	201,000
	3,32	,830	201	201,000
	4,33	15,175	201	201,000
	15,79	32,761	201	201,000
Петров П.П. Сколько Вам лет (укажите полное число лет)? Какое у Вас образование? Как за последний год изменилась Ваша жизнь и ее отдельные стороны? - Ваш личный доход По Вашему мнению, как работают следующие органы власти: хорошо или плохо? - Президент	40,21	14,129	86	86,000
	3,37	,841	86	86,000
	4,09	14,744	86	86,000
	11,37	27,593	86	86,000
Сидоров С.С. Сколько Вам лет (укажите полное число лет)? Какое у Вас образование? Как за последний год изменилась Ваша жизнь и ее отдельные стороны? - Ваш личный доход По Вашему мнению, как работают следующие органы власти: хорошо или плохо? - Президент	46,60	18,305	225	225,000
	3,41	,831	225	225,000
	2,02	6,524	225	225,000
	7,20	22,295	225	225,000
Всего	45,23	16,974	512	512,000
	3,37	,831	512	512,000
	3,28	12,091	512	512,000
	11,27	27,927	512	512,000

- Средние значения таких дискриминантных переменных, как **возраст респондента** и **изменение личного дохода** различны в каждой из трех исследуемых групп.
- Также видны различия в мнениях о **качестве работы Президента**.
- Далее проводим тест на равенство групповых средних и проверяем отсутствие корреляции.

## Лабораторная работа №19

- **Статистика Лямбда Уилкса** показывает, что респонденты значимо делятся на группы по двум критериям: возраст и мнение о качестве работы президента. Однако дискриминирующая способность переменных оказалась слабой, так как значение Лямбды Уилкса близко к 1. Это повлияет на качество прогноза. Чтобы добиться лучшего результата необходимо использовать больше переменных в анализе.

Критерии равенства групповых средних

	Лямбда Уилкса	F	ст.св.1	ст.св.2	Знач.
Сколько Вам лет (укажите полное число лет)?	,982	4,694	2	509	,010
Какое у Вас образование?	,997	,692	2	509	,501
Как за последний год изменилась Ваша жизнь и ее отдельные стороны? - Ваш личный доход	,991	2,185	2	509	,114
По Вашему мнению, как работают следующие органы власти: хорошо или плохо? - Президент	,980	5,110	2	509	,006



## Лабораторная работа №19

Так как в анализе деление производится на три классификационные группы, то дискриминантных функций будет две. Обе дискриминантные функции дают значимые результаты для разделения между группами, т.к. 0,002 и 0,038 < 0,05. Однако **D1** объясняет больший % дисперсии, чем **D2**. Следовательно, для прогноза лучше использовать **D1**.

Лямбда Уилкса

Критерий для функций	Лямбда Уилкса	Хи-квадрат	ст.св.	Знач.
От 1 до 2	,954	23,809	8	,002
2	,984	8,411	3	,038

Собств. знач

Функция	Собственное значение	% дисперсии	Суммарный %	Каноническая корреляция
1	,031 <sup>а</sup>	64,8	64,8	,173
2	,017 <sup>а</sup>	35,2	100,0	,128

а. Для анализа использовались первые 2 из канонических дискриминантных функций.

Коэффициенты канонической дискриминантной функции

	Функция	
	1	2
Сколько Вам лет (укажите полное число лет)?	-,024	,054
Какое у Вас образование?	-,354	-,013
Как за последний год изменилась Ваша жизнь и ее отдельные стороны? - Ваш личный доход	,038	,007
По Вашему мнению, как работают следующие органы власти: хорошо или плохо? - Президент (Константа)	,025	,016
	1,864	-2,596

Нестандартизованные коэффициенты

Итоговые дискриминантные функции выглядят следующим образом:

$$D1 = 1,864 - 0,024x - 0,354y + 0,038z + 0,025t$$

$$D2 = -2,596 + 0,054x - 0,013y + 0,007z + 0,016t$$

где:

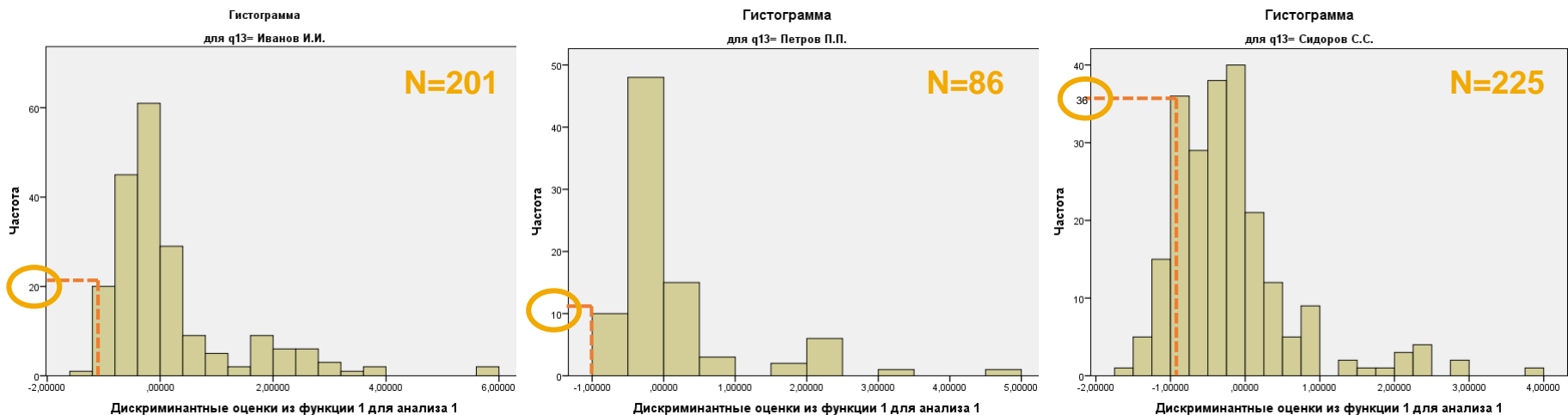
- x – Возраст респондента;
- y – Образование респондента;
- z – «Как изменилась ваша жизнь...? - Личный доход»;
- t – «По Вашему мнению, как работают...? - Президент»

## Лабораторная работа №19

На основе дискриминантной модели, подставив соответствующие параметры в функцию, **можно сделать прогноз** о том, за кого из кандидатов проголосует большинство респондентов.

Полученное значение **D** необходимо найти на диаграммах распределения значений дискриминантной функции.

Для вычисления прогноза воспользуемся функцией  **$D1 = 1,864 - 0,024x - 0,354y + 0,038z + 0,025t$**



Например,  **$D1 = -1$**

Голосовали за Иванова 201 чел. из выборки, а с таким значением функции – 20 чел., за Петрова – 86 и 10 чел. соответственно, за Сидорова – 225 и 36 чел. соответственно. Следовательно, респонденты с таким значением дискриминантной функции с большей вероятностью будут голосовать за Сидорова. В SPSS выводится таблица статистик для прогнозирования каждого отдельного случая (см. Лекцию 10, слайд 34).

# Определение качества модели

Результаты классификации<sup>а</sup>

		Если бы выборы состоялись в следующее воскресенье, за кого из представленных кандидатов Вы бы, скорее всего, проголосовали?	Предсказанная принадлежность к группе			Всего
			Иванов И.И.	Петров П.П.	Сидоров С.С.	
Исходный	Количество	Иванов И.И.	42	67	92	201
		Петров П.П.	17	46	23	86
		Сидоров С.С.	23	87	115	225
		Несгруппированные наблюдения	205	192	223	620
%		Иванов И.И.	20,9	33,3	45,8	100,0
		Петров П.П.	19,8	53,5	26,7	100,0
		Сидоров С.С.	10,2	38,7	51,1	100,0
		Несгруппированные наблюдения	33,1	31,0	36,0	100,0

а. 39,6% исходных сгруппированных наблюдений классифицированы правильно.

- Показывает % верно классифицированных случаев.
- Точность прогнозов невысокая и составляет 39,6%.
- За кандидата Сидорова С.С. проголосует больше всего респондентов (453 чел.).
- Однако для более корректного анализа лучше использовать больше наблюдений.

# Лабораторная работа №20

## **Дискриминантный анализ: пошаговый метод**

## Лабораторная работа №20

---

1. Откройте файл **Elections.sav**.
2. Спрогнозируйте, за кого из кандидатов проголосует большинство респондентов (переменная q13) пошаговым методом. При этом учтите, что многие пока не решили, за кого будут голосовать, затруднялись/уклонялись от ответа на вопрос.
3. Интерпретируйте полученные результаты.

## Лабораторная работа №20

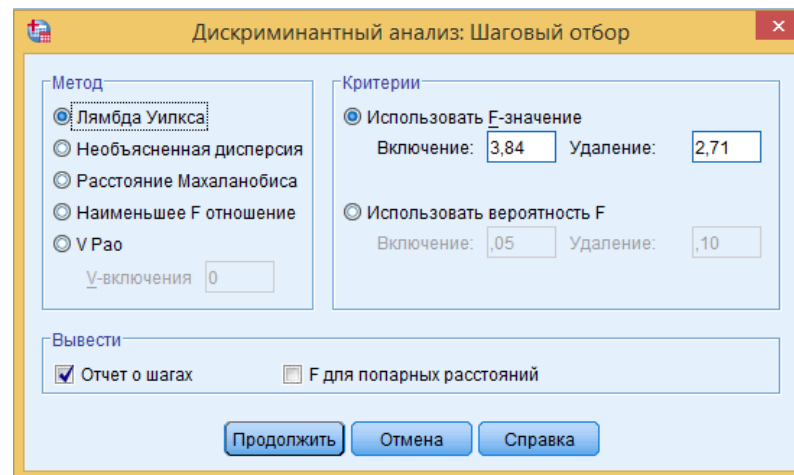
---

1. Команды **«Анализ»** → **«Классификация»** → **«Дискриминантный анализ»**.
2. Зависимая переменная q13 («Если бы выборы...?») переносится в поле **«Группировать по»**.
3. Задайте **«Диапазон»** от 1 до 3.
4. В качестве независимых переменных (факторных признаков) возьмем s1 (Возраст), s2 (Образование), q6\_2 (Как за последний год изменилась ваша жизнь? – Личный доход) и q8\_1 (По вашему мнению, как работают следующие органы власти: хорошо или плохо? – Президент) и перенесем в поле **«Независимые переменные»**.
5. Выберите **«Пошаговый метод»**.

## Лабораторная работа №20

---

6. Во вкладке «**Статистики**» отметьте:
  - в разделе «**Описательные статистики**»: «**Средние**», «**Однофакторный дисперсионный анализ**».
  - в разделе «**Коэффициенты функции**»: «**Нестандартизированные**».
  - в разделе «**Матрицы**»: «**Внутригрупповая корреляция**».
7. Во вкладке «**Метод**» выберите «**Лямбда Уилкса**»
8. Во вкладке «**Классифицировать**» отмечаем:
  - «**Все группы равны**»
  - «**Внутригрупповая**»
  - вывод «**Итоговой таблицы**»
  - «**Отдельно по группам**»



## Лабораторная работа №20

- Согласно критерию Лямбда Уилкса обе дискриминантные функции дают значимые результаты для разделения между группами, т.к.  $0,001$  и  $0,004 < 0,05$
- Согласно спрогнозированным данным за Сидорова С.С. проголосует больше всего респондентов (539 чел.).
- Точность прогноза 38,1%.
- Таким образом, оба метода дают схожие результаты.

Лямбда Уилкса

Критерий для функций	Лямбда Уилкса	Хи-квадрат	ст.св.	Знач.
От 1 до 2	,963	19,330	4	,001
2	,984	8,300	1	,004

Результаты классификации<sup>а</sup>

		Если бы выборы состоялись в следующее воскресенье, за кого из представленных кандидатов Вы бы, скорее всего, проголосовали?	Предсказанная принадлежность к группе			Всего
			Иванов И.И.	Петров П.П.	Сидоров С.С.	
Исходный	Количество	Иванов И.И.	28	69	104	201
		Петров П.П.	8	46	32	86
		Сидоров С.С.	13	91	121	225
		Несгруппированные наблюдения	142	196	282	620
	%	Иванов И.И.	13,9	34,3	51,7	100,0
		Петров П.П.	9,3	53,5	37,2	100,0
		Сидоров С.С.	5,8	40,4	53,8	100,0
		Несгруппированные наблюдения	22,9	31,6	45,5	100,0

<sup>а</sup> 38,1% исходных сгруппированных наблюдений классифицированы правильно.