



# Метод главных компонент. Теория и практика на примере выявления политических предпочтений

**Филатов А.Ю.**

Институт систем энергетики им.Л.А.Мелентьева,  
Иркутский государственный университет

[http://polnolunie.baikal.ru/me/mat\\_ec.htm](http://polnolunie.baikal.ru/me/mat_ec.htm),

<http://polnolunie.baikal.ru/me/metrix.htm>,

<http://matec.isu.ru>,

[http://fial\\_.livejournal.com](http://fial_.livejournal.com)



# Снижение размерности признакового пространства

$$x^{(1)}, \dots, x^{(p)} \rightarrow z^{(1)}, \dots, z^{(p')}, p' < p$$

## Причины:

1. Сжатие объемов хранимой информации.
2. Наглядное представление данных ( $p = 1, 2, 3$ ).
3. Лаконизм моделей, упрощение счета и интерпретации.
4. Увеличение точности выводов, зависящей от  $n / (p+1)$ .
5. Борьба с мультиколлинеарностью – взаимозависимостью регрессоров.

## Требования к новым показателям:

1. Максимальная информативность.
2. Взаимная некоррелированность.
3. Минимальное искажение геометрической структуры исходных данных.

## Ситуации, в которых снижение размерности осуществить легко:

1. Дублирование информации (исключение).
2. Наличие неинформативных переменных (исключение).
3. Наличие однотипных переменных (агрегирование)



# Метод главных компонент

## 1. Подготовительный этап

- 1) Центрирование и нормирование переменных – переход к  $(x_i^{(j)} - \bar{x}^{(j)}) / \sqrt{\sigma_j}$
- 2) Вычисление матрицы ковариаций

$$\Sigma = \begin{pmatrix} \hat{\sigma}_{11} & \dots & \hat{\sigma}_{1p} \\ \dots & \dots & \dots \\ \hat{\sigma}_{p1} & \dots & \hat{\sigma}_{pp} \end{pmatrix}, \quad \hat{\sigma}_{kj} = \frac{1}{n} \sum_{i=1}^n (x_i^{(k)} - \bar{x}^{(k)}) (x_i^{(j)} - \bar{x}^{(j)}) =$$
$$= \text{КОВАР}(x_1^{(k)}, \dots, x_n^{(k)}; x_1^{(j)}, \dots, x_n^{(j)}).$$

## 2. Решение характеристического уравнения $|\Sigma - \lambda E| = 0$

- 1) Нахождение собственных чисел  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p'} > 0$
- 2) Нахождение собственного вектора  $l^{(k)}$  для каждого корня  $\lambda_k$   
 $(\Sigma - \lambda_k E)l^{(k)} = 0, \quad \|l^{(k)}\| = 1.$

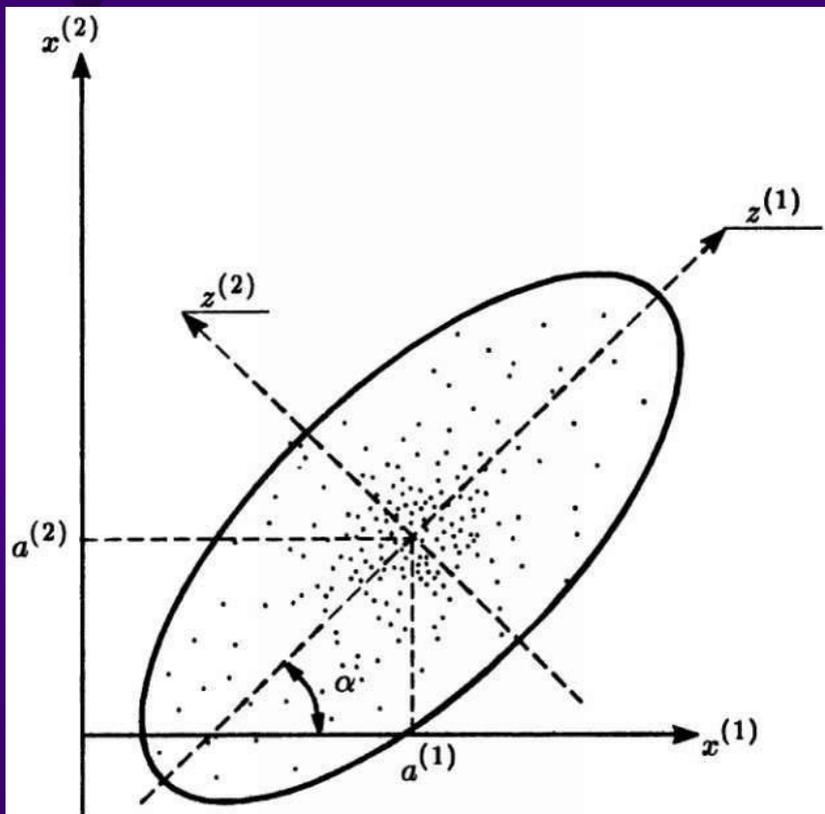
## 3. Переход к новым переменным $Z = XL$

$z^{(k)} = Xl^{(k)}, \quad k = 1, \dots, p'$  – новые переменные, «главные компоненты»

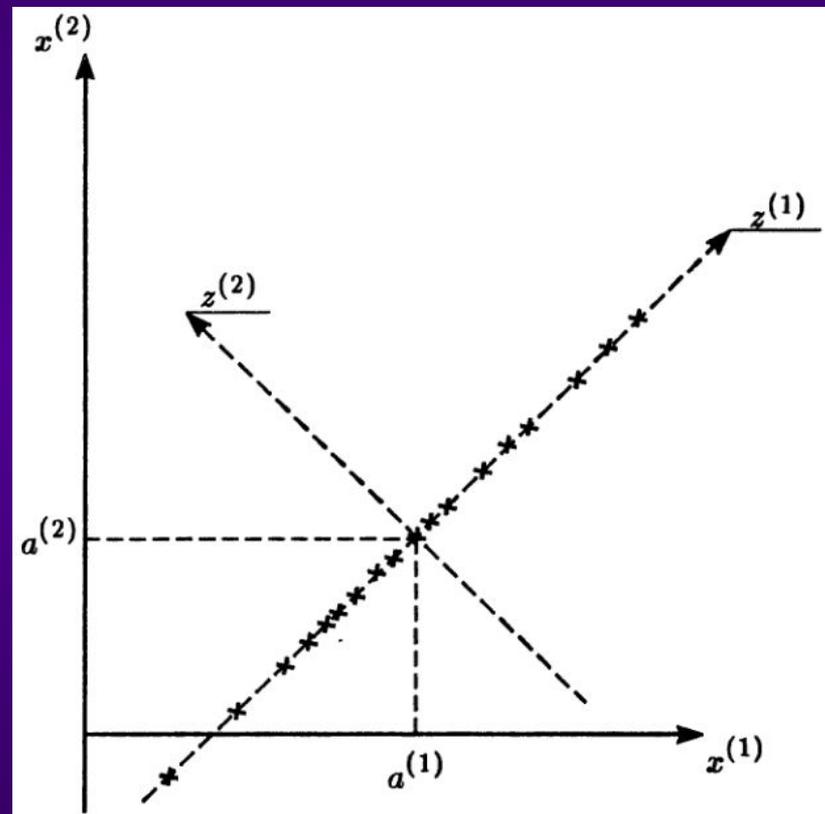
$l_{p'} = \frac{\lambda_1 + \dots + \lambda_{p'}}{\lambda_1 + \dots + \lambda_p}$  – доля дисперсии, вносимая первыми  $p'$  главными компонентами



# Геометрическая демонстрация метода главных компонент



**Рис.1.** Умеренный разброс  
точек вдоль  $z^{(2)}$



**Рис.2.** Вырожденный случай:  
отсутствие разброса вдоль  $z^{(2)}$



# Проблема интерпретации главных компонент

**Матрица нагрузок главных компонент на исходные переменные:**

$$A \in R^{p \times p'}, \quad A = L\Lambda^{1/2}, \quad \Lambda^{1/2} = \text{diag}\{\sqrt{\lambda_j}\}, \quad a_{ij} = r(x^{(i)}, z^{(j)})$$

## Наблюдения – ежемесячные

данные  $x^{(1)}$  – число торговых точек, где распространяется

$x^{(2)}$  продукция, шт.

$x^{(3)}$  – расходы на рекламу, руб.

$x^{(4)}$  – доля новинок в ассортименте, %

$x^{(5)}$  – средний месячный доход на душу населения, руб.

– количество праздников, шт.

$$A = \begin{pmatrix} z^{(1)} & z^{(2)} \\ 0,95* & -0,19 \\ 0,97* & -0,17 \\ 0,94* & -0,28 \\ 0,24 & 0,88* \\ 0,56 & 0,67* \end{pmatrix} \begin{matrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \\ x^{(4)} \\ x^{(5)} \end{matrix}$$

$$\sum_{i=1}^p a_{ij}^2 = a_{1j}^2 + a_{2j}^2 + \dots + a_{pj}^2 = \lambda_j$$

$$\sum_{j=1}^p a_{ij}^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 = 1$$

$z^{(1)}$  тесно связана с  $x^{(1)}, x^{(2)}, x^{(3)}$ .

$z^{(2)}$  тесно связана с  $x^{(4)}, x^{(5)}$ .



# Эмпирическое исследование: Опрос ВЦИОМ, осень 2007

**Дано:** 40 понятий

Каждый из 1589 респондентов выбирает несколько (в пределах 15), вызывающих у него положительную или отрицательную реакцию

## **Дополнительные вопросы:**

1. Намерение голосовать за ту или иную партию на предстоящих выборах
2. Демографические характеристики (пол, возраст, образование, доход)
3. Заинтересованность политикой
4. Степень влияния на жизнь в стране
5. Регион и тип населенного пункта, где проживает респондент
6. Доверие к Президенту, другим органам власти

## **Модификация данных:**

Каждому понятию присвоено значение

- 1, если оно вызывает у респондента отрицательные чувства,
- 1, если оно вызывает у респондента положительные чувства,
- 0, если оно не вызывает никаких чувств.



# Исходные данные

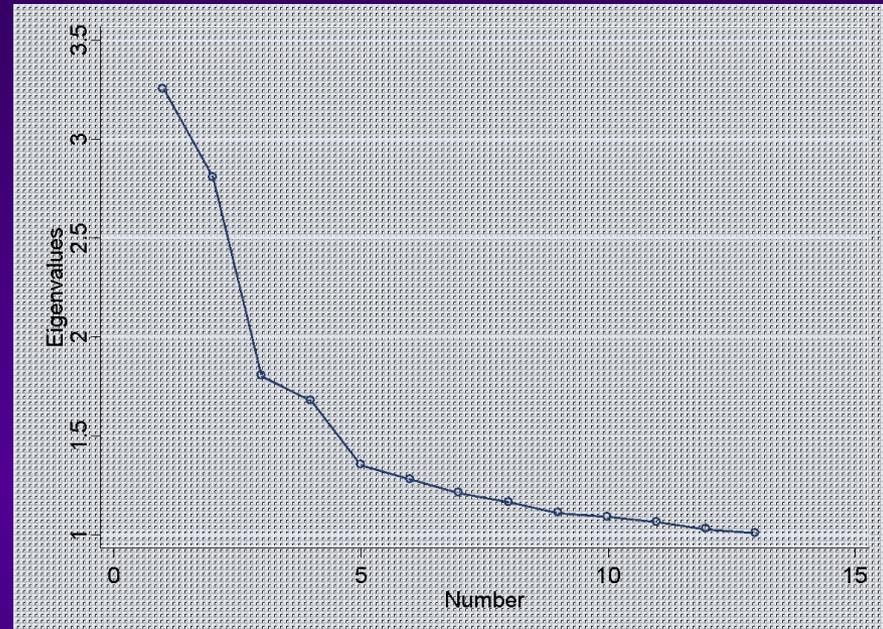
**Табл.1.** Доля респондентов, оценивших понятие как положительное или отрицательное

Понятие	Полож	Отриц
01. Нация	0,21	0,08
02. Порядок	0,57	0,01
03. Свобода	0,37	0,03
04. Рынок	0,10	0,15
05. Русские	0,34	0,02
06. Запад	0,02	0,23
07. Социализм	0,11	0,11
08. Коммунизм	0,07	0,19
09. Демократия	0,15	0,09
10. Традиция	0,29	0,01
11. Патриотизм	0,34	0,01
12. Государство	0,26	0,03
13. Конкурентоспособн.	0,05	0,07
14. Суверенитет	0,07	0,05
15. Элита	0,02	0,41
16. Партия	0,02	0,16
17. Власть	0,09	0,18
18. Справедливость	0,49	0,02
19. Оппозиция	0,01	0,17
20. Бизнес	0,07	0,13

Понятие	Полож	Отриц
21. СССР	0,12	0,08
22. Церковь	0,21	0,02
23. Революция	0,01	0,22
24. Собственность	0,14	0,04
25. Успех	0,31	0,00
26. Либерализм	0,01	0,14
27. Реформа	0,06	0,14
28. Стабильность	0,38	0,00
29. Труд	0,31	0,00
30. Индивидуализм	0,02	0,12
31. Нерусские	0,02	0,29
32. Равенство	0,18	0,02
33. Коллективизм	0,06	0,09
34. Мораль	0,22	0,03
35. Права человека	0,32	0,02
36. Богатство	0,12	0,01
37. Россия	0,28	0,00
38. Достаток	0,37	0,01
39. Прогресс	0,21	0,01
40. Капитализм	0,15	0,02

# Интерпретация главных компонент

**Рис.3.** Собственные числа главных компонент



## $z^{(1)}$ – «толерантность»

Высокое значение соответствует отсутствию отрицательной реакции на слова «элита», «нерусские», «рынок», «запад», «власть» и «реформа», а также отсутствию положительной реакции на слова «справедливость» и «труд».

## $z^{(2)}$ – «экономическая свобода»

Высокое значение соответствует положительной реакции на слова «свобода», «бизнес», «успех», «богатство», «достаток», «прогресс» и «капитализм» и отрицательной реакции на «социализм», «коммунизм», «СССР», «революцию» и «коллективизм».



# Матрица факторных нагрузок

Табл.2. Коэффициенты корреляции главных компонент и исходных переменных

Понятие	$z^{(1)}$	$z^{(2)}$
01. Нация	0,11	-0,08
02. Порядок	-0,18	0,01
03. Свобода	-0,13	0,20
04. Рынок	0,26	0,08
05. Русские	-0,15	0,03
06. Запад	0,21	0,10
07. Социализм	-0,13	-0,28
08. Коммунизм	0,05	-0,32
09. Демократия	0,11	0,07
10. Традиция	-0,06	-0,04
11. Патриотизм	-0,14	-0,15
12. Государство	-0,17	-0,03
13. Конкурентоспособн.	0,07	0,12
14. Суверенитет	-0,08	0,01
15. Элита	0,30	0,04
16. Партия	0,04	-0,14
17. Власть	0,26	-0,09
18. Справедливость	-0,30	0,02
19. Оппозиция	0,12	-0,06
20. Бизнес	0,17	0,27

Понятие	$z^{(1)}$	$z^{(2)}$
21. СССР	-0,01	-0,34
22. Церковь	-0,13	-0,01
23. Революция	0,13	-0,26
24. Собственность	0,13	0,14
25. Успех	-0,16	0,21
26. Либерализм	0,15	-0,01
27. Реформа	0,23	-0,02
28. Стабильность	-0,16	0,00
29. Труд	-0,26	-0,08
30. Индивидуализм	0,05	0,10
31. Нерусские	0,25	-0,12
32. Равенство	-0,18	0,06
33. Коллективизм	0,02	-0,22
34. Мораль	-0,05	-0,07
35. Права человека	-0,15	0,12
36. Богатство	-0,15	0,25
37. Россия	-0,03	0,07
38. Достаток	0,11	0,25
39. Прогресс	-0,03	0,27
40. Капитализм	-0,09	0,22



# Распределение предпочтений – все респонденты

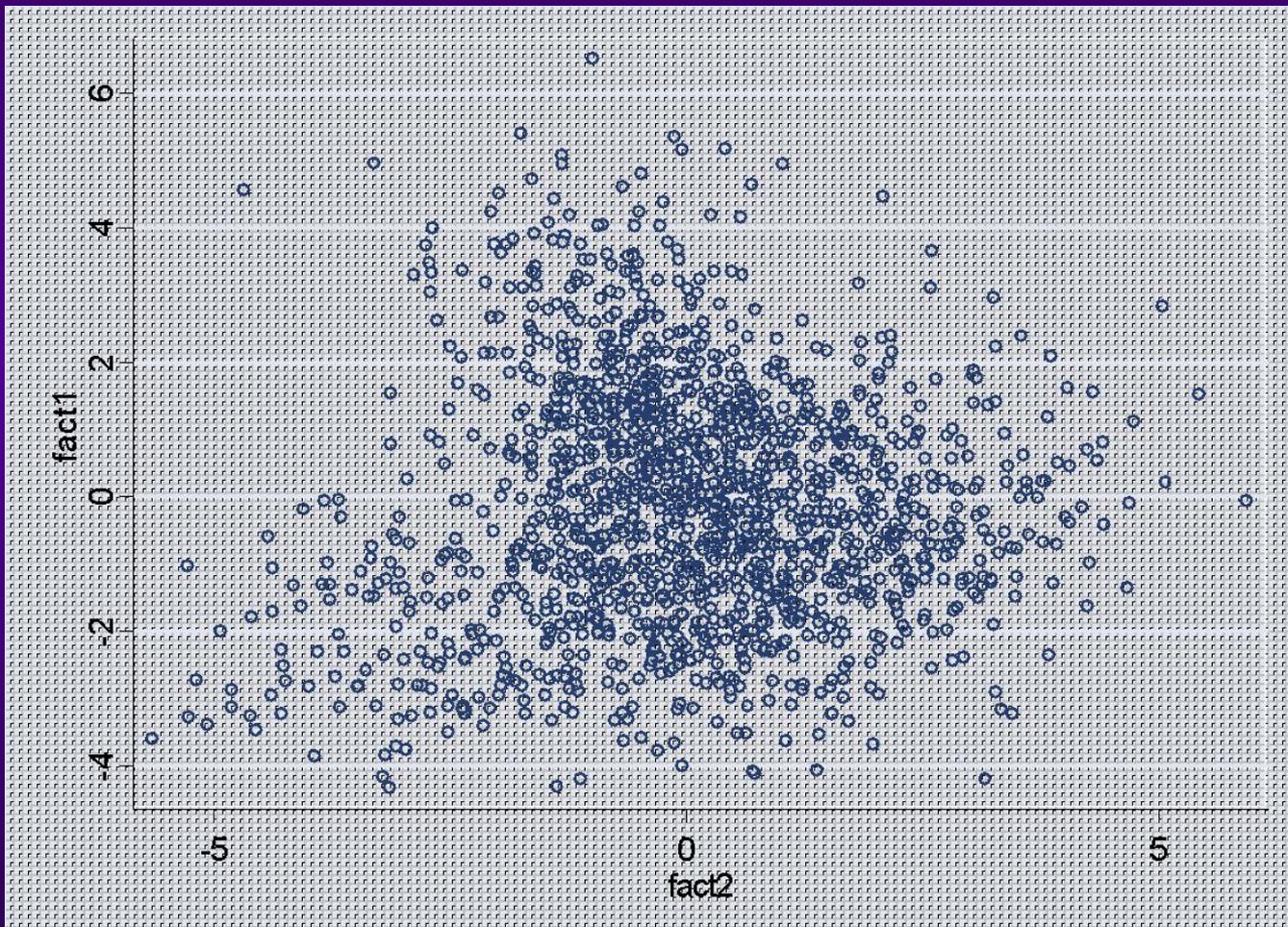


Рис.4. Распределение предпочтений – все респонденты



# 1. Партийные симпатии

Предпочтения	%	$z^{(1)}$	$z^{(2)}$
1. Аграрная	0,63	-0,16	-0,92
2. ЕР	45,72	0,05	0,30
3. КПРФ	7,12	-0,76	-1,59
4. ЛДПР	4,22	-0,53	0,69
6. Патриоты России	0,25	0,22	-0,10
7. Справедливая Россия	6,17	-0,60	-0,87
8. Свободная Россия	0,69	-0,43	0,31
9. СПС	0,57	-0,47	1,14
10. Яблоко	0,76	-0,56	0,20
11. Республиканская	0,25	-0,16	1,36
13. Демократическая	0,19	-0,25	0,75
14. Не голосовать	17,88	0,23	-0,06

Табл.3. Партийные симпатии

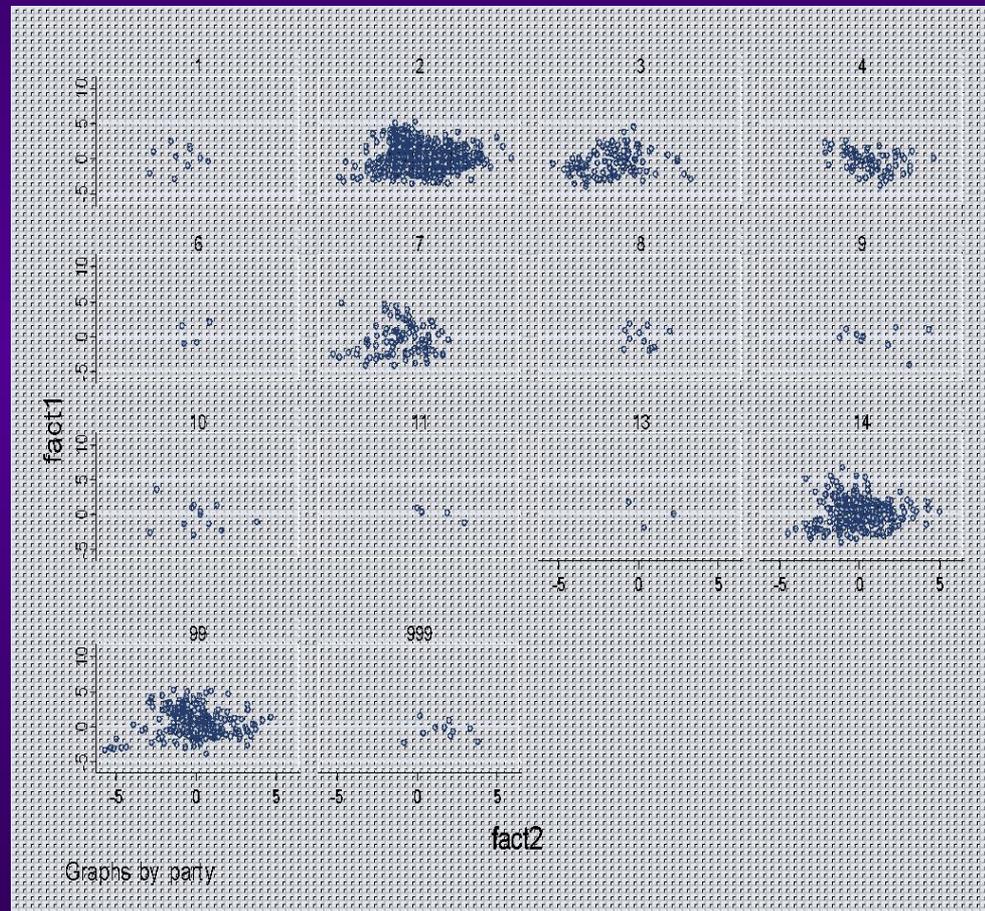


Рис.5. Партийные симпатии



## 2. Доход

Доход	Чел.	$z^{(1)}$	$z^{(2)}$
1. Высокий	41	-0,1	-0,16
2. Выше среднего	126	0,82	0,28
3. Средний	989	-0,04	0,19
4. Ниже среднего	383	0,15	-0,48
5. Низкий	79	-0,05	-0,59

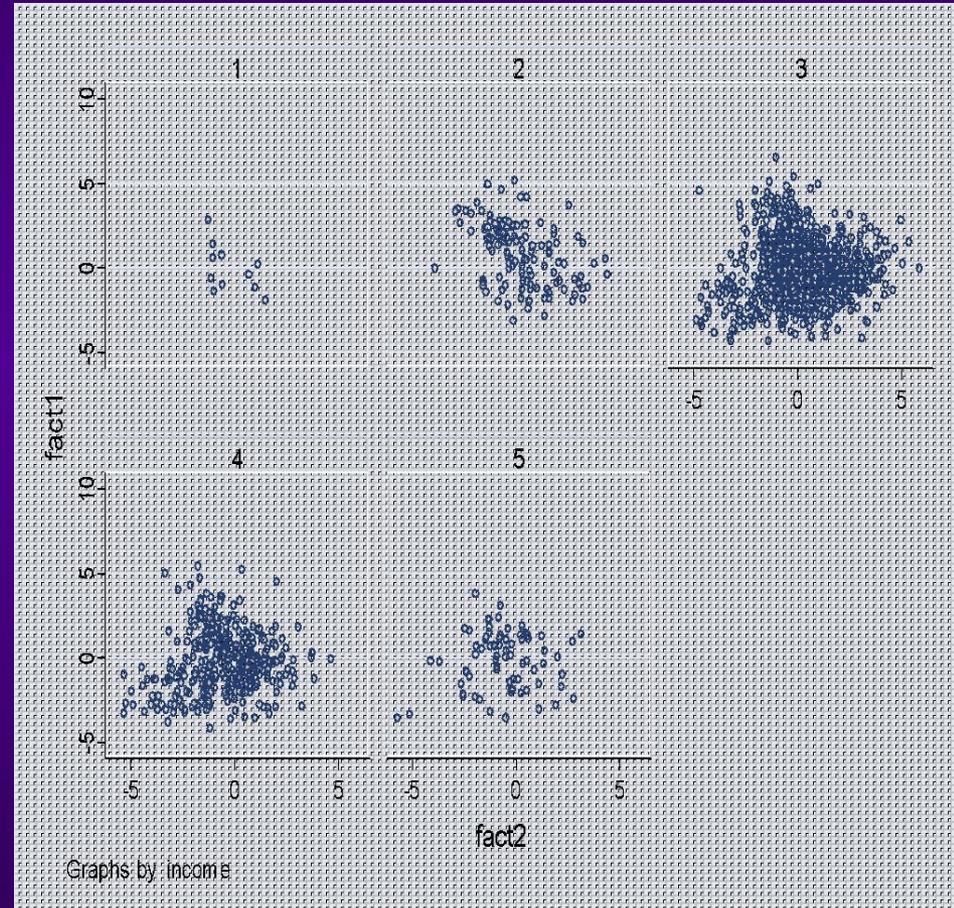


Табл.4. Доход

Рис.6. Доход



### 3. Образование

Образование	Чел.	$z^{(1)}$	$z^{(2)}$
1. Начальное	42	-0,24	-1,38
2.	118	-0,03	-0,80
3.	538	-0,06	0,04
4.	545	-0,17	0,12
5.	77	0,34	0,40
6. Высшее	268	0,18	0,11

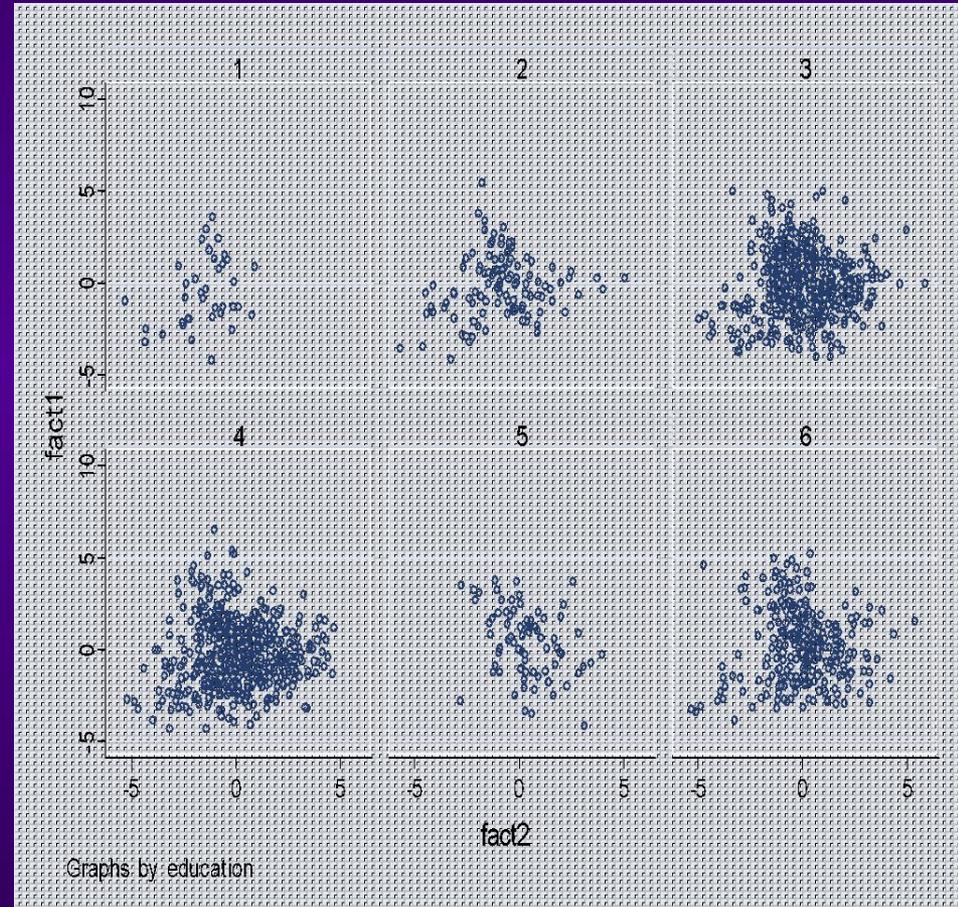


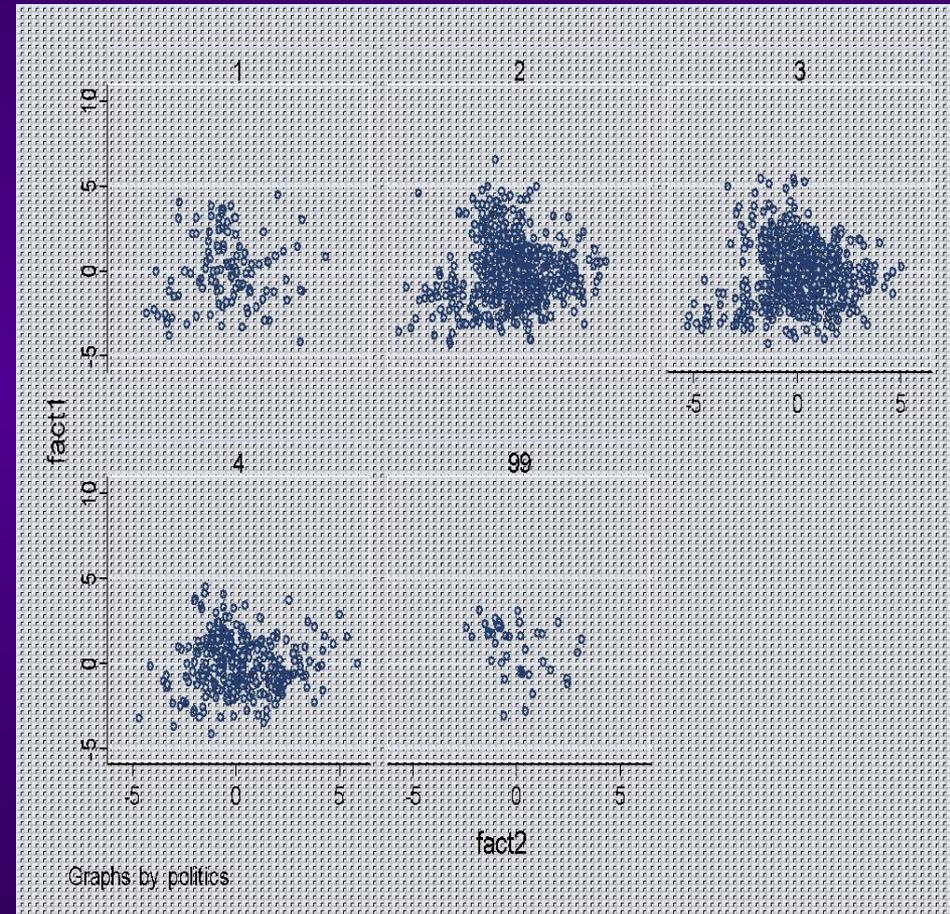
Табл.5. Образование

Рис.7. Образование

## 4. Заинтересованность политикой

Заинтерес. политикой	Чел.	$z^{(1)}$	$z^{(2)}$
1. Высокая	112	0,37	-0,34
2.	559	0,02	-0,16
3.	556	-0,15	0,13
4. Совсем не интересна	309	0,02	0,21
5. Затрудн. ответить	52	1,11	-0,21

Если наложить данные графики на графики партийных предпочтений, обнаружим, что наиболее интересуются политикой сторонники **КПРФ** и **СР** (низкие значения обоих факторов), а наименее – сторонники **правых партий** (высокое значение второго фактора).



**Табл.6.** Заинтересованность политикой

**Рис.8.** Заинтересованность политикой



## 5. Федеральный округ

Федеральный округ	Чел.	$z^{(1)}$	$z^{(2)}$
1. Центральный	418	0,32	0,02
2. Северо-Западный	154	0,06	-0,08
3. Южный	253	0,21	-0,18
4. Поволжский	343	-0,24	-0,08
5. Уральский	92	0,31	0,32
6. Сибирский	210	-0,57	0,21
7. Дальневосточный	118	-0,18	0,02

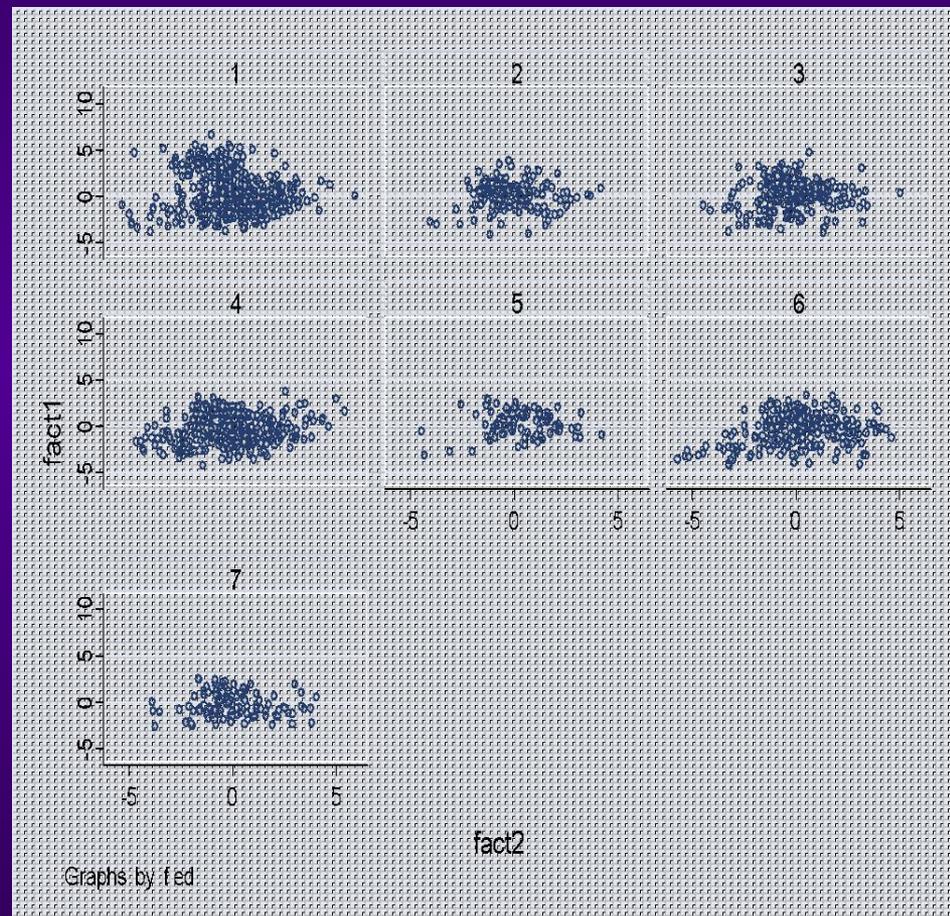
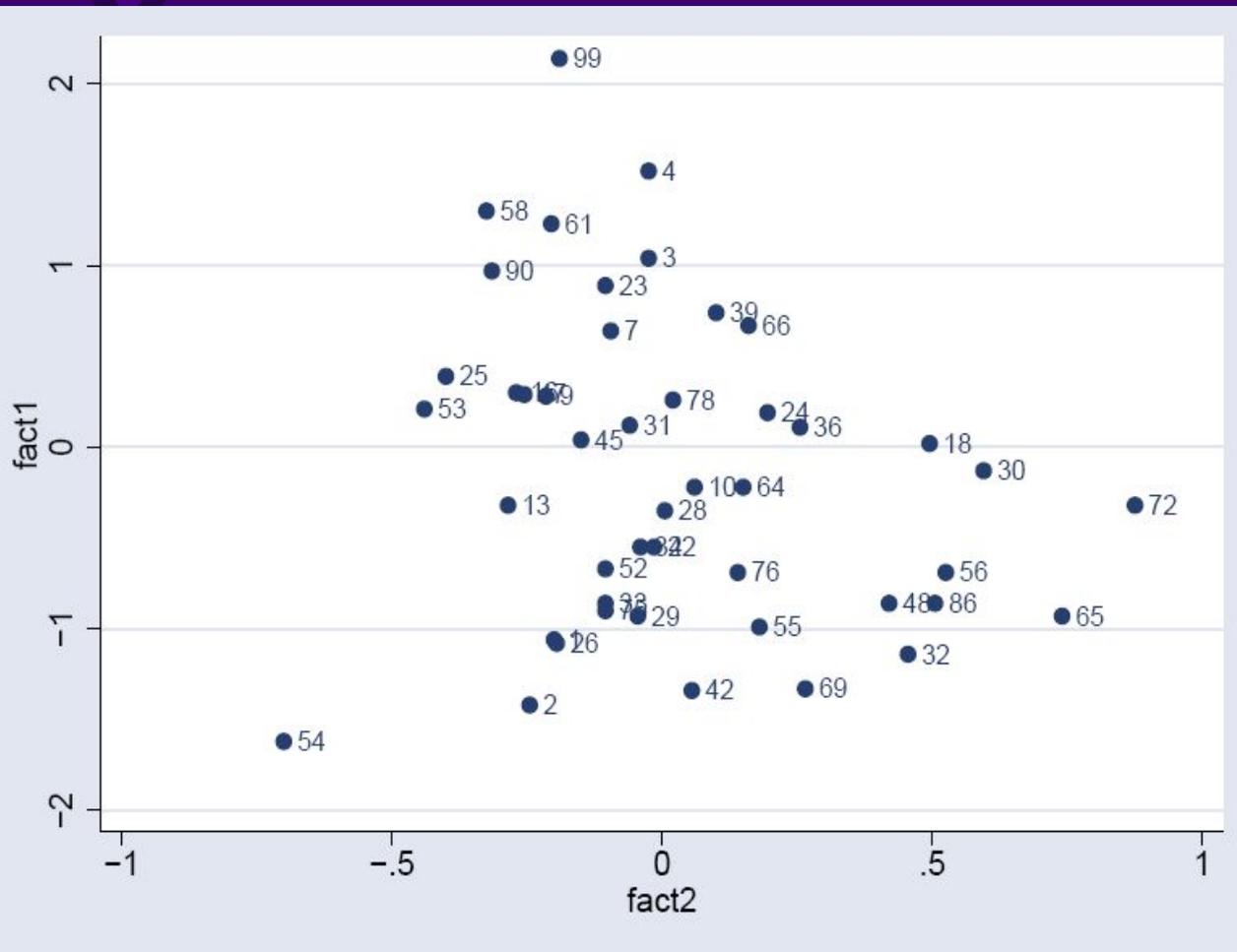


Табл.7. Федеральный округ

Рис.9. Федеральный округ

## 6. Регионы (обратные координаты!)



- 99 – Москва
- 04 – Алтай
- 58 – Пензенская обл.
- 61 – Ростовская обл.
- 54 – Новосибирская обл.
- 02 – Башкирия
- 42 – Кемеровская обл.
- 69 – Тверская обл.
- 53 – Новгородская обл.
- 25 – Приморский край
- 72 – Тюменская обл.
- 65 – Сахалинская обл.
- 30 – Астраханская обл.

Рис.10. Регионы



## 7. Размер населенного пункта

Размер насел. пункта	Чел.	$z^{(1)}$	$z^{(2)}$
1. Москва, СП	163	1,54	-0,25
2.	139	-0,44	-0,33
3.	148	0,01	0,14
4.	290	-0,34	0,32
5.	135	0,23	-0,35
6.	171	0,39	0,61
7.	116	-0,18	0,10
8. Село	426	-0,28	-0,17

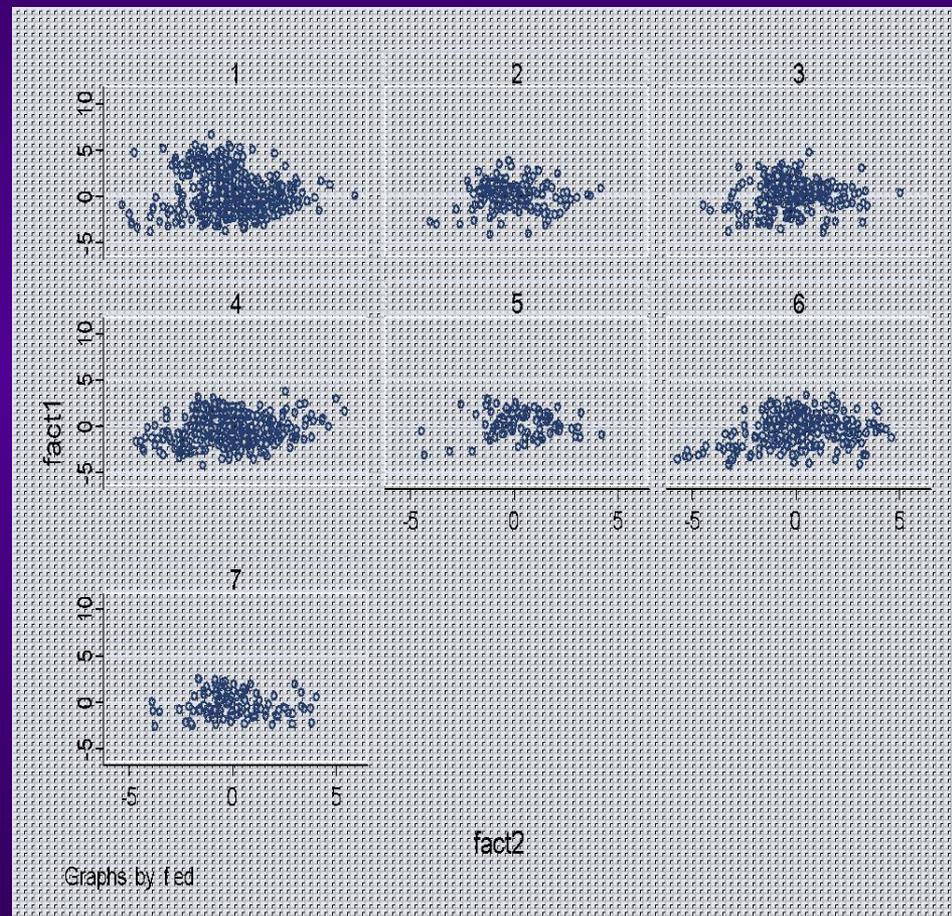
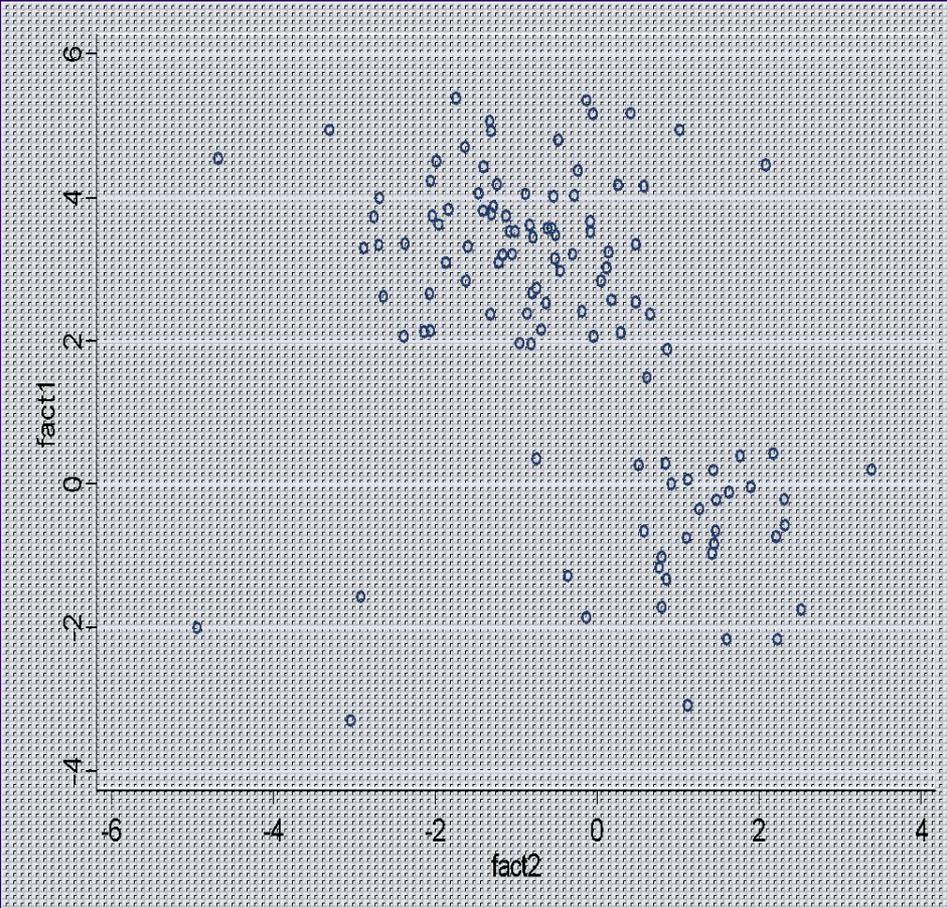


Табл.8. Размер населенного пункта

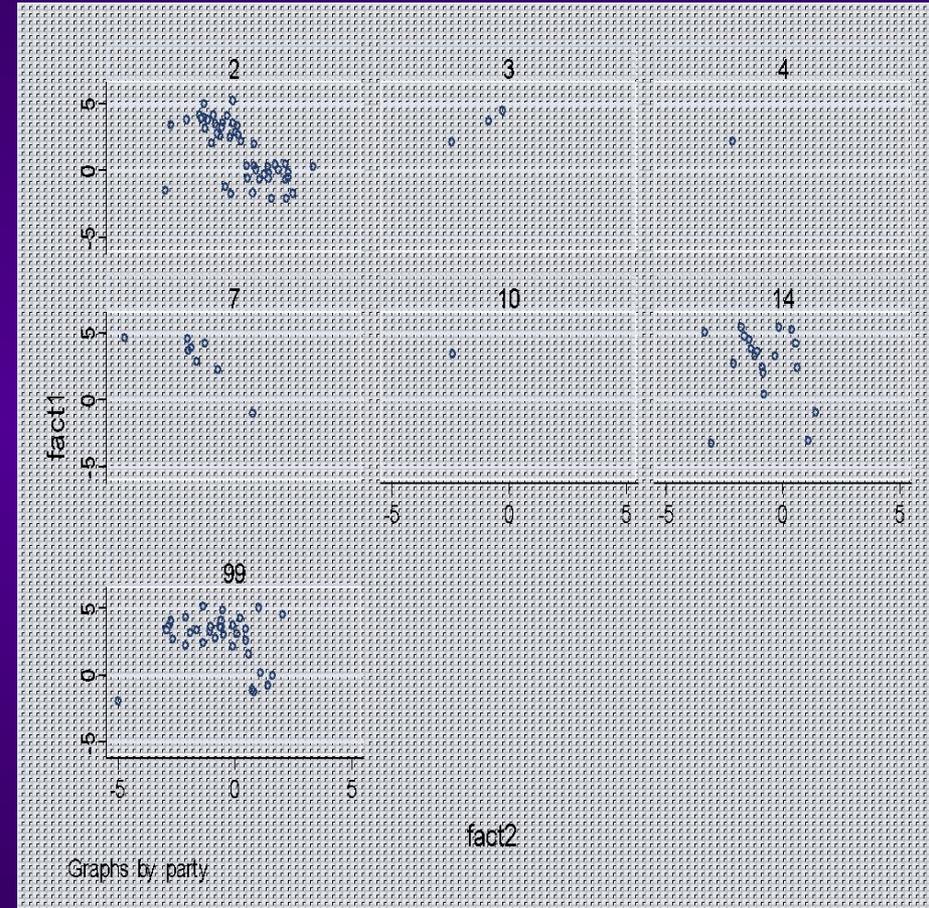
Рис.11. Размер населенного пункта



## 8. Москва



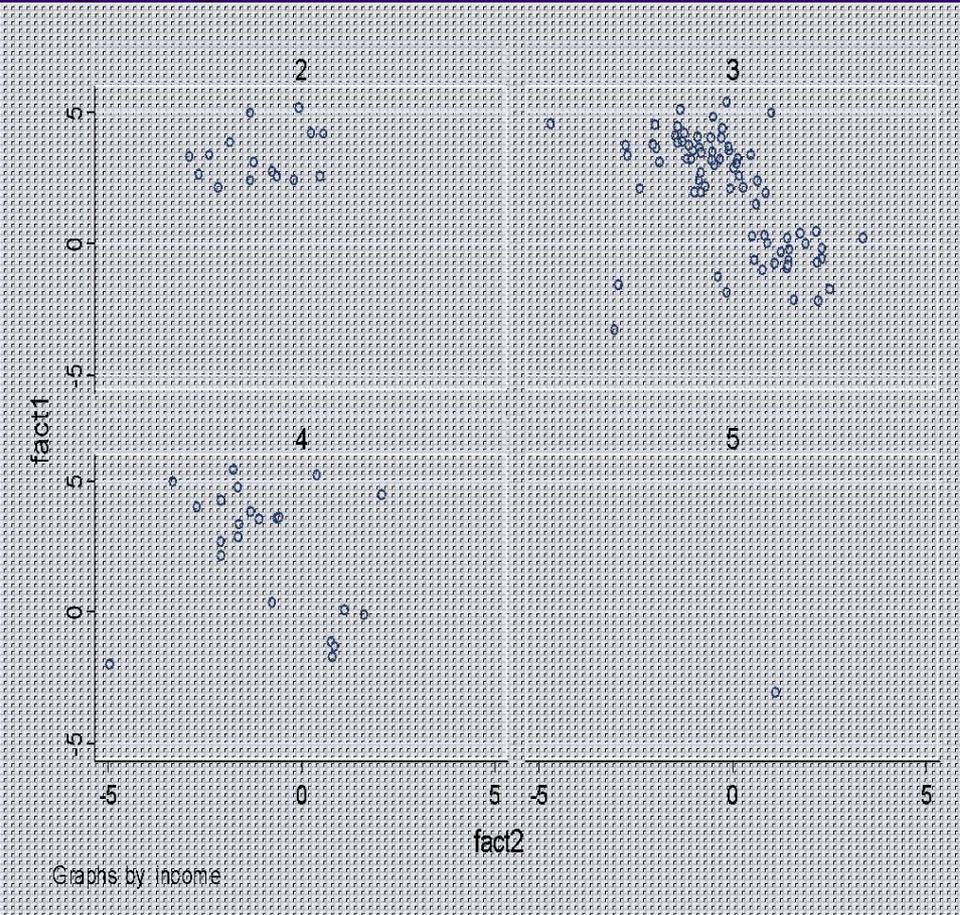
**Рис.12.** Москва. 2 кластера:  
недовольные «экономически продвинутые»  
и довольные «непродвинутые»



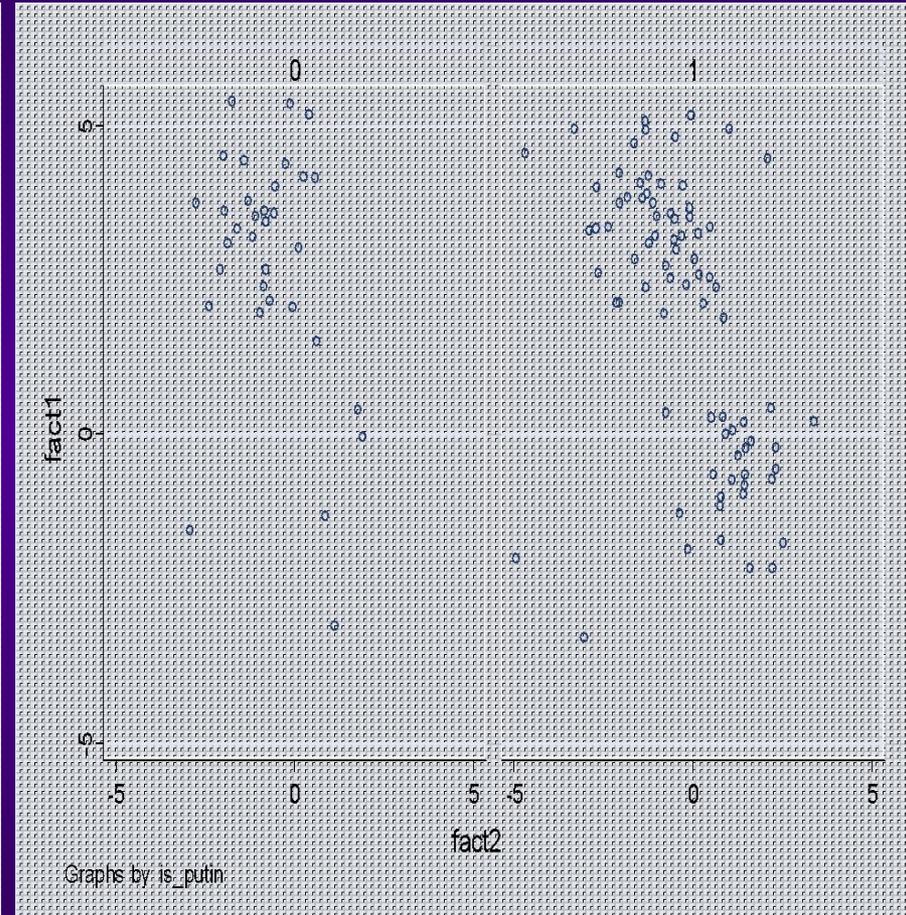
**Рис.13.** Москва. Партии  
2 – ЕР, 3 – КПРФ, 4 – ЛДПР,  
7 – СР, 10 – Яблоко, 14 – не голосов.



## 8. Москва



**Рис.14.** Москва. Доход  
2 – выше среднего, 5 – бедные



**Рис.15.** Москва. Доверие к Путину  
0 – не доверяют, 1 – доверяют



## 9. Доверие к Путину. Готовность к акциям протеста

Доверие к Путину	Чел.	$z^{(1)}$	$z^{(2)}$
1. Да	1252	-0,02	0,11
2. Нет	201	0,02	-0,44
3. Затрудн. ответить	134	0,18	-0,39

Табл.9. Доверие к Путину

Готовность к акциям протеста	Чел.	$z^{(1)}$	$z^{(2)}$
1. Да	347	-0,23	-0,36
2. Нет	1074	0,08	0,15
3. Затрудн. ответить	165	-0,01	-0,25

Табл.10. Готовность к акциям протеста

## Модель множественного выбора

Образование (0→1) Не влияет

Доверие к власти (0→1) ЕР (+)

Жизнь в деревне (0/1)

Бедность (0→1) ЛДПР(+)

Возраст (в годах) КПРФ (+), СР (+)

Пол ЛДПР (М)

Степень влияния на жизнь в стране

$$u_{ij} = \theta_{0j} + \theta_{1j}x_i^{(1)} + \dots + \theta_{pj}x_i^{(p)} + \beta_1(v_{i1} - y_{j1})^2 + \beta_2(v_{i2} - y_{j2})^2$$



# Игра: моделирование избирателя

**Табл.11.** Процент голосующих за партии (мужчины / женщины)

$z^{(1)}$	$z^{(2)}$	ЕР	КПРФ	ЛДПР	СР
0	0	73% / 86%	7% / 4%	11% / 2%	9% / 8%
3,4	0	83% / 92%	4% / 2%	7% / 1%	6% / 4%
-3,4	0	56% / 76%	13% / 8%	15% / 3%	14% / 13%
0	3,4	78% / 94%	1% / 1%	17% / 3%	3% / 3%
0	-3,4	45% / 61%	31% / 20%	4% / 1%	19% / 18%

**Табл.12.** Процент голосующих за партии (мужчины)

Путин	Влиян.	Бедн.	Село	Возраст	ЕР	КПРФ	ЛДПР	СР
1	1	1	1	30	94%	1%	1%	3%
1	1	4	0	30	78%	2%	16%	4%
0	0	1	1	30	61%	15%	13%	11%
0	0	4	0	30	23%	10%	62%	5%
1	1	1	1	60	85%	5%	1%	9%
1	1	4	0	60	74%	8%	8%	10%
0	0	1	1	60	37%	38%	4%	20%
0	0	4	0	60	21%	36%	29%	14%



# Игра: моделирование избирателя

## Профили репрезентативных избирателей ключевых партий:

**ЕР (97%):** женщина,  $z^{(1)}=0$ ,  $z^{(2)}=0$ , 30 лет, город, обеспеченная, доверяет Путину, доверяет Думе, влияет на жизнь.

**ЛДПР (88%):** мужчина,  $z^{(1)}=-1,7$ ,  $z^{(2)}=1,7$ , 30 лет, город, бедный, не доверяет Путину, доверяет Думе, не влияет на жизнь.

**КПРФ (33%):** мужчина,  $z^{(1)}=-1,7$ ,  $z^{(2)}=-1,7$ , 60 лет, село, бедный, не доверяет Путину, не доверяет Думе, не влияет на жизнь.

**Не голосует:** низкое образование, низкий доход, недоверие, молодежь, город.

**Табл.13.** Процент голосующих за партии (факт / при наличии «против всех»)

Путин	ЕР	КПРФ	ЛДПР	СР	Против всех
Ноябрь, 2007	72% / 56%	11% / 9%	7% / 5%	10% / 8%	/ 22%
50%	61% / 35%	16% / 13%	11% / 8%	12% / 8%	/ 36%
0%	43% / 18%	25% / 16%	19% / 11%	12% / 6%	/ 49%

## Идеи дальнейших исследований:

1. Позиции партий – сознательное поведение (равновесие Нэша).
2. Изменение экономического положения за последние годы (экон.голосование).



*Спасибо  
за внимание!*