Эконометрика-1

Филатов Александр Юрьевич

(Главный научный сотрудник, доцент ШЭМ ДВФУ)

alexander.filatov@gmail.com

http://vk.com/alexander.filatov, http://vk.com/baikalreadings

Практика-3

Регрессионный анализ. МНК. Метод главных компонент

Линейная регрессия

Задача 1а «Продажи в зависимости от цены»

Продавец настольных игр собрал статистику по ценам и объемам продаж игры «Диксит» с января 2016 по апрель 2018 г.

– 0,040	162,7
0,023	49,9
0,105	17,30

= ЛИНЕЙН ($y_1,\ldots,y_n$; $x_1^{(1)},\ldots,x_n^{(p)}$; 1; 1).

$3 \times (p+1)$ ⇒ формула ⇒ Ctrl-Shift-Enter

	объем	цена
	y	$x^{(1)}$
янв.16	91	1990
фев.16	93	1990
мар.16	84	1990
апр.16	77	1990
май.16	69	2190
июн.16	49	2190
июл.16	53	2190
авг.16	55	2190
сен.16	62	2190
окт.16	69	2190
ноя.16	68	2190
дек.16	109	2190
…	…	…
апр.18	72	2290

$$\hat{y}_i = \underset{(49,9)}{162,7} - \underset{(0,023)}{0,040}\, x_i^{(1)}, \qquad \hat{R}^2 = 0,105 = 10,5\%$$

$$t_0 = \frac{162,7}{49,9} = 3,26, \qquad t_1 = \frac{-0,040}{0,023} = -1,74$$

$t_{\text{крит}}$ = СТЬЮДРАСПОБР(0,05; 28 – 1 – 1) = 2,06.

Цена не значима при α = 0,05.

$$\theta_0 \in [162,7 - 2,06 \cdot 49,9; 162,7 + 2,06 \cdot 49,9] = [60,1; 265,3],$$

$$\theta_1 \in [-0,040 - 2,06 \cdot 0,023; -0,040 + 2,06 \cdot 0,023] = [-0,087; 0,007].$$

$$F_{\text{эмп}} = \frac{0,105}{1-0,105} \cdot \frac{26}{1} = 3,04 < 4,23 = F_{0,05}(1; 26) = F_{\text{крит}}.$$

Модель не значима при α = 0,05.

Линейная регрессия

Задача 1b «Продажи в зависимости от цены и рекламы»

Продавец также собрал данные о рекламном бюджете фирмы.

	объем	цена	рекл
	y	$x^{(1)}$	$x^{(2)}$
янв.16	91	1990	10
фев.16	93	1990	30
мар.16	84	1990	30
апр.16	77	1990	10
май.16	69	2190	10
июн.16	49	2190	0
июл.16	53	2190	0
авг.16	55	2190	20
сен.16	62	2190	20
окт.16	69	2190	20
ноя.16	68	2190	20
дек.16	109	2190	20
…	…	…	…
апр.18	72	2290	20

0,470	-0,039	149,2
0,170	0,020	44,8
0,315	15,43	#Н/Д

$$\hat{y}_i = \underset{(44,8)}{149,2} - \underset{(0,020)}{0,039}\, x_i^{(1)} + \underset{(0,170)}{0,470}\, x_i^{(2)}, \qquad \hat{R}^2 = 0,315 = 31,5\%$$

$$t_0 = 3,33, \quad t_1 = -1,89, \quad t_2 = 2,77$$

$t_{\text{крит}}$ = СТЬЮДРАСПОБР(0,05; 28 – 2 – 1) = 2,06.

Цена не значима, реклама значима при α = 0,05.

$$\theta_0 \in [56,9; 241,5], \quad \theta_1 \in [-0,081; 0,004], \quad \theta_2 \in [0,120; 0,819].$$

$$F_{\text{эмп}} = \frac{0,315}{1-0,315} \cdot \frac{25}{2} = 5,74 > 3,39 = F_{0,05}(2; 25) = F_{\text{крит}}.$$

Модель значима при α = 0,05.

Линейная регрессия

Задача 1с «Продажи в зависимости от цены, рекламы и праздников»

Продавец учел число праздничных дней в месяце.

	объем	цена	рекл	празд
	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$
янв.16	91	1990	10	6
фев.16	93	1990	30	1
мар.16	84	1990	30	2
апр.16	77	1990	10	0
май.16	69	2190	10	3
июн.16	49	2190	0	1
июл.16	53	2190	0	0
авг.16	55	2190	20	0
сен.16	62	2190	20	0
окт.16	69	2190	20	0
ноя.16	68	2190	20	1
дек.16	109	2190	20	0
…	…	…	…	…
апр.18	72	2290	20	0

2,70	0,471	-0,045	158,8
1,62	0,164	0,020	43,7
0,386	14,91	#Н/Д	#Н/Д

$$\hat{y}_i = \underset{(43,7)}{158,8} - \underset{(0,020)}{0,045}\, x_i^{(1)} + \underset{(0,164)}{0,471}\, x_i^{(2)} + \underset{(1,62)}{2,70}\, x_i^{(3)},$$

$$\hat{R}^2 = 0,386 = 38,6\%.$$

$$t_0 = 3,64,\ \ t_1 = -2,22,\ \ t_2 = 2,87,\ \ t_3 = 1,67.$$

$t_{\text{крит}}$ = СТЬЮДРАСПОБР(0,05; 28 – 3 – 1) = 2,06.

Цена и реклама значимы, праздники нет.

$$\theta_0 \in [68,7;\ 249,0],\quad \theta_1 \in [-0,086;\ -0,003],$$

$$\theta_2 \in [0,132;\ 0,809],\quad \theta_3 \in [-0,64;\ 6,04].$$

$$F_{\text{эмп}} = \frac{0,386}{1-0,386}\cdot\frac{24}{3} = 5,03 > 3,01 = F_{0,05}(3;\ 24) = F_{\text{крит}}.$$

Модель значима при α = 0,05.

Линейная регрессия

Задача 1d «Продажи в зависимости от цены, рекламы, праздников и цены ближайшего конкурента»

	объем	цена	рекл	празд	конк
	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$
янв.16	91	1990	10	6	1990
фев.16	93	1990	30	1	1990
мар.16	84	1990	30	2	1990
апр.16	77	1990	10	0	1990
май.16	69	2190	10	3	2190
июн.16	49	2190	0	1	2190
июл.16	53	2190	0	0	2190
авг.16	55	2190	20	0	2190
сен.16	62	2190	20	0	2190
окт.16	69	2190	20	0	2190
ноя.16	68	2190	20	1	2190
дек.16	109	2190	20	0	2390
…	…	…	…	…	…
апр.18	72	2290	20	0	2290

0,111	4,22	0,623	-0,177	201,3
0,022	1,17	0,118	0,029	31,6
0,713	10,41	#Н/Д	#Н/Д	#Н/Д

$$\hat{y}_i = \underset{(31,6)}{201,3} - \underset{(0,029)}{0,177}\, x_i^{(1)} + \underset{(0,118)}{0,623}\, x_i^{(2)} + \underset{(1,17)}{4,22}\, x_i^{(3)} + \underset{(0,022)}{0,111}\, x_i^{(4)},$$

$$\hat{R}^2 = 0,713 = 71,3\%.$$

$$t_0 = 6,37,\ t_1 = -6,02,\ t_2 = 5,27,\ t_3 = 3,61,\ t_4 = 5,12.$$

$t_{\text{крит}}$ = СТЬЮДРАСПОБР(0,05; 23) = 2,07.

Все переменные значимы при α = 0,05.

$$F_{\text{эмп}} = 14,30 > 2,80 = F_{\text{крит}}.$$

Модель значима при α = 0,05.

Линейная регрессия

Задача 1е «Продажи в зависимости от цены, рекламы, праздников, цены ближайшего конкурента и будущих праздников»

	объем	цена	рекл	празд	конк	пр+1
	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$
янв.16	91	1990	10	6	1990	1
фев.16	93	1990	30	1	1990	2
мар.16	84	1990	30	2	1990	0
апр.16	77	1990	10	0	1990	3
май.16	69	2190	10	3	2190	1
июн.16	49	2190	0	1	2190	0
июл.16	53	2190	0	0	2190	0
авг.16	55	2190	20	0	2190	0
сен.16	62	2190	20	0	2190	0
окт.16	69	2190	20	0	2190	1
ноя.16	68	2190	20	1	2190	0
дек.16	109	2190	20	0	2390	5
…	…	…	…	…	…	…
апр.18	72	2290	20	0	2290	3

5,29	0,085	4,31	0,641	-0,142	173,3
0,77	0,013	0,68	0,068	0,018	18,7
0,908	6,02	#Н/Д	#Н/Д	#Н/Д	#Н/Д

$$\hat{y}_i = \underset{(18,7)}{173,3} - \underset{(0,018)}{0,142}\, x_i^{(1)} + \underset{(0,068)}{0,641}\, x_i^{(2)} + \underset{(0,68)}{4,31}\, x_i^{(3)} + \underset{(0,013)}{0,085}\, x_i^{(4)} + \underset{(0,77)}{5,29}\, x_i^{(5)}, \quad \hat{R}^2 = 0,908 = 90,8\%.$$

$$t_0 = 9,26,\ t_1 = -7,96,\ t_2 = 9,36,\ t_3 = 6,38,$$

$$t_4 = 6,51,\ t_5 = 6,84.$$

Все переменные значимы при $\alpha = 0,05$.

$$F_{\text{эмп}} = 43,58 > 2,66 = F_{\text{крит}}.$$

Модель значима при $\alpha = 0,05$.

Возможные дальнейшие шаги: тренд, доход, макропоказатели, учет инфляции (через индексацию), нелинейные зависимости. **Они могут ухудшить модель!**

Метод главных компонент

Задача 2 «Эмпирическое исследование на основе опроса ВЦИОМ перед выборами 2007 г.» (Алексей Захаров, НИУ ВШЭ)

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 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

Матрица факторных нагрузок

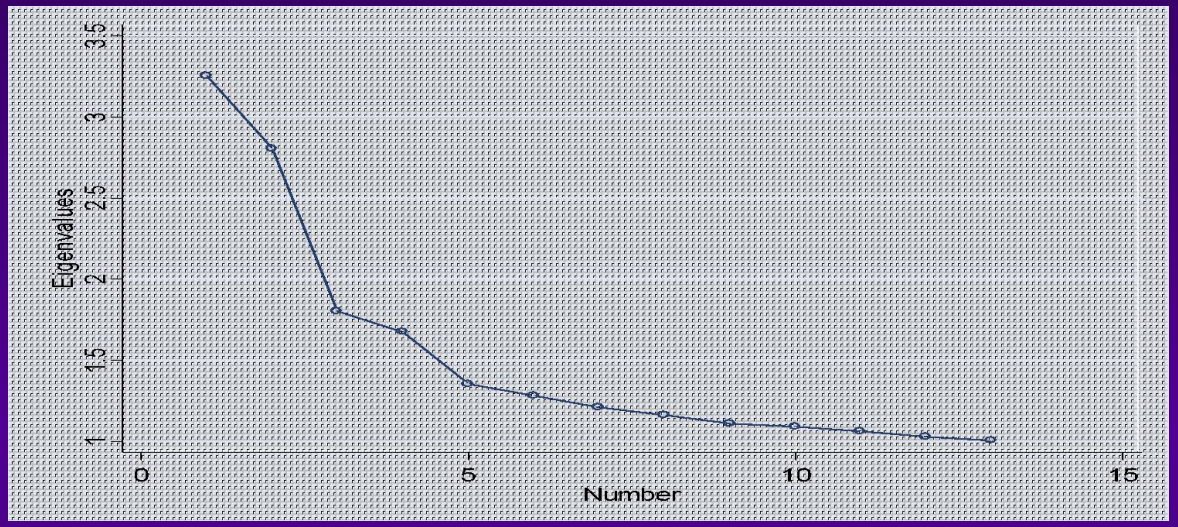
Табл.2. Коэффициенты корреляции главных компонент и исходных переменных

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

Интерпретация главных компонент

Рис.1. Собственные числа главных компонент



$z^{(1)}$ – «толерантность» / «успешность» / «безразличие»

Высокое значение соответствует отсутствию отрицательной реакции на слова «элита», «нерусские», «рынок», «запад», «власть» и «реформа», и 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 от рицательной реакции на «социализм», «коммунизм», «СССР», «революцию» и «коллективизм».

Распределение предпочтений – все респонденты

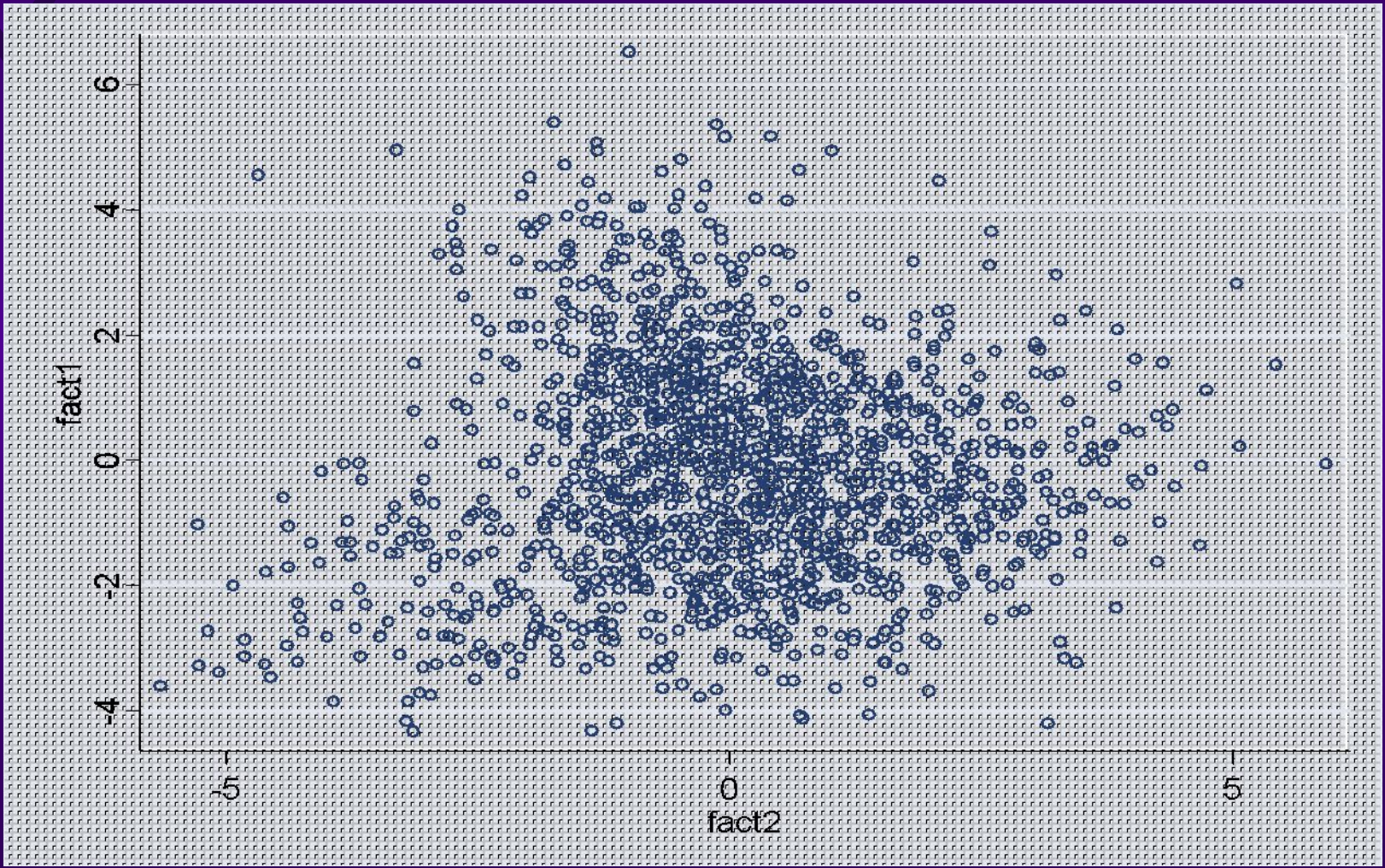


Рис.2. Распределение предпочтений – все респонденты

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

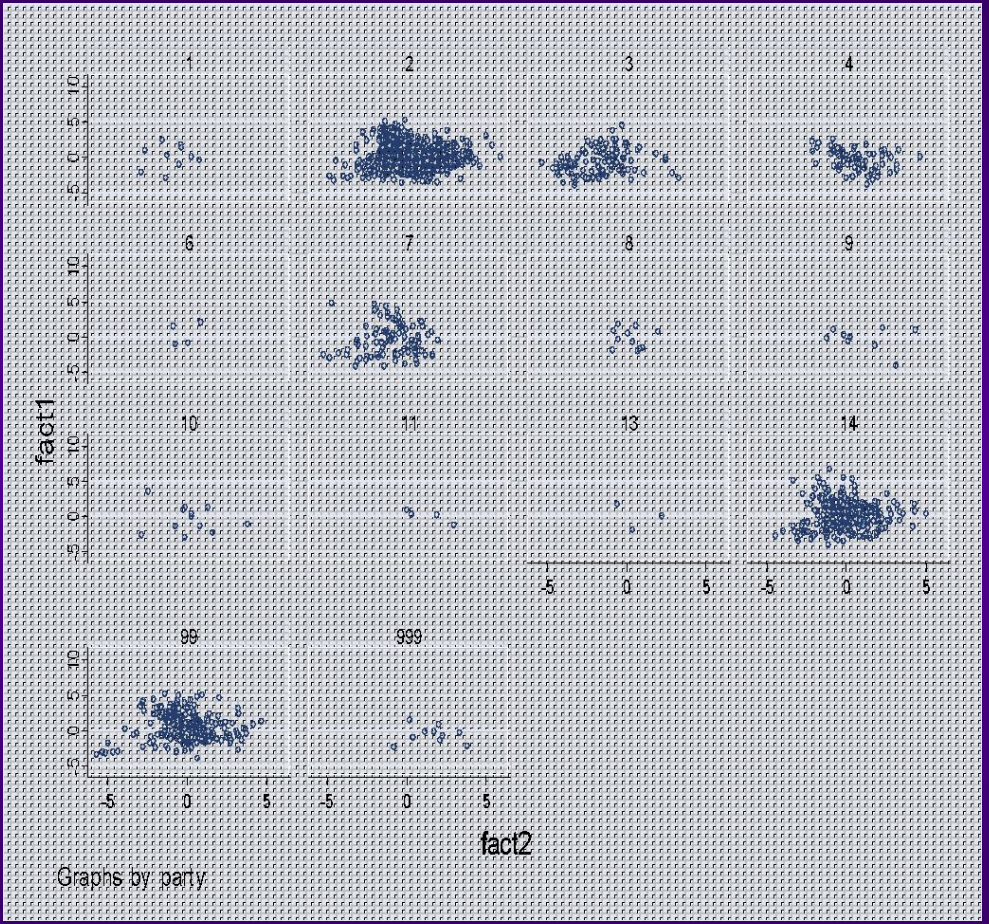


Рис.3. Партийные симпатии

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

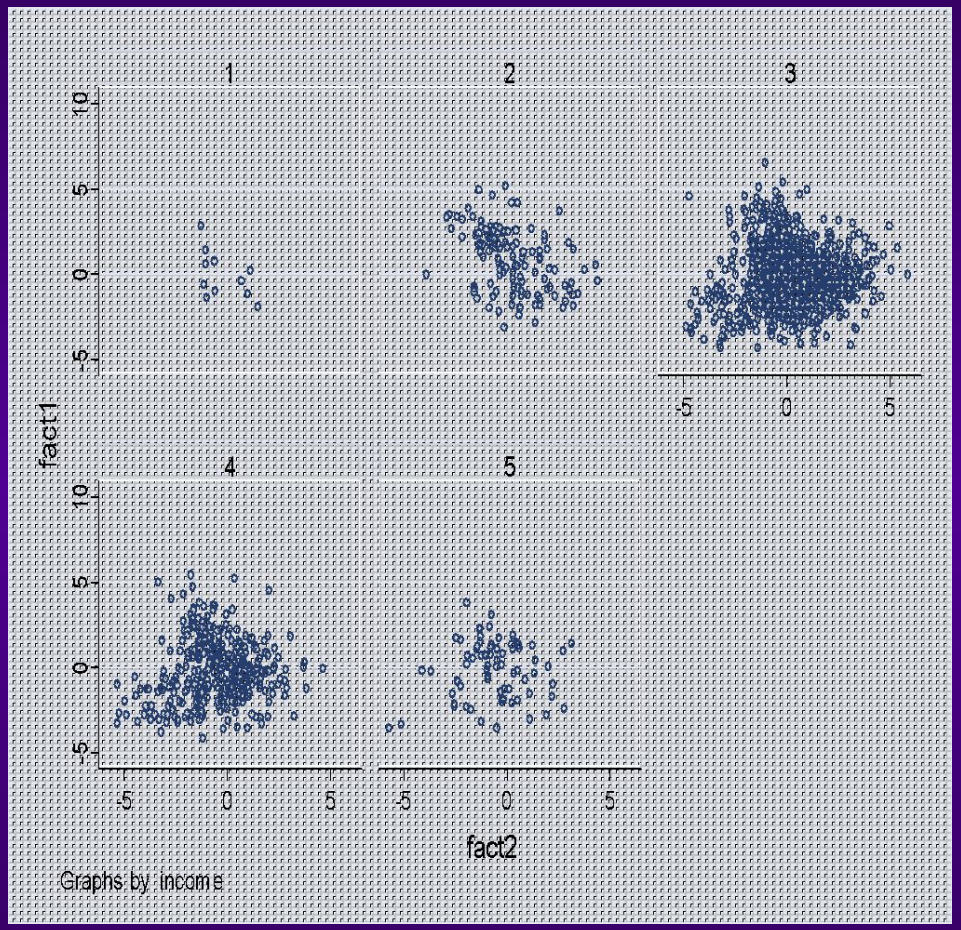


Рис.4. Доход

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

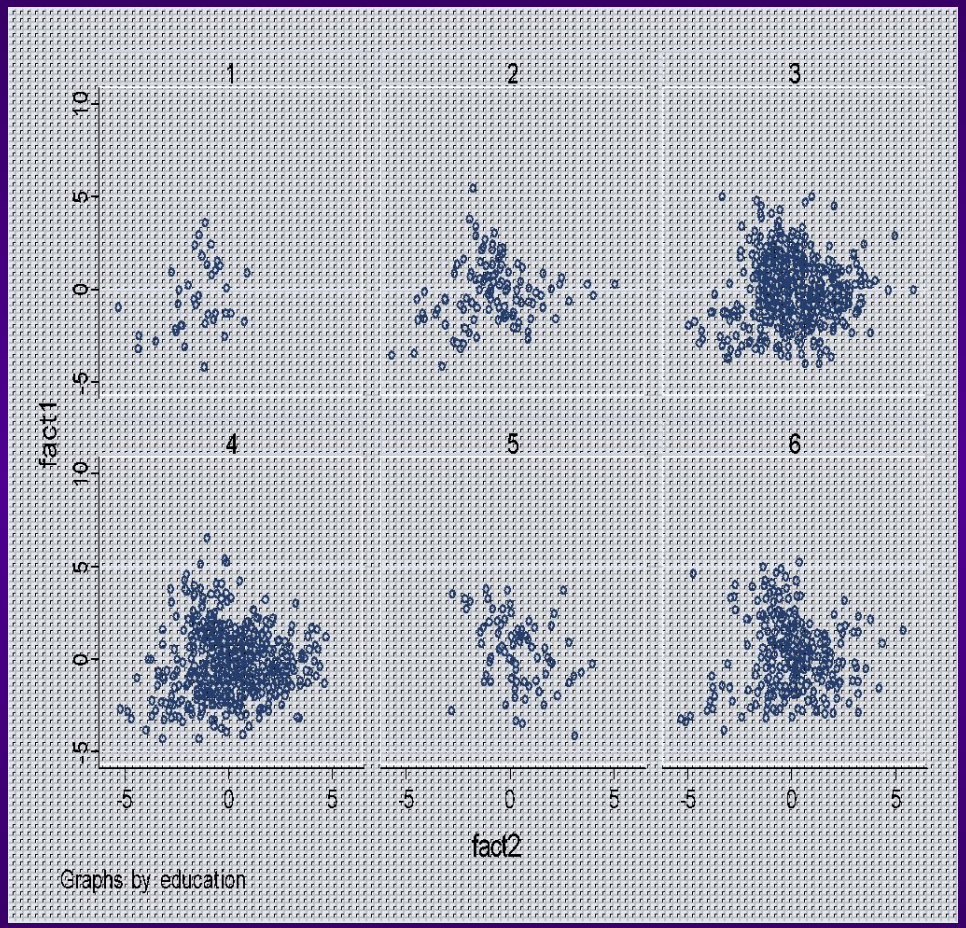


Рис.5. Образование

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).

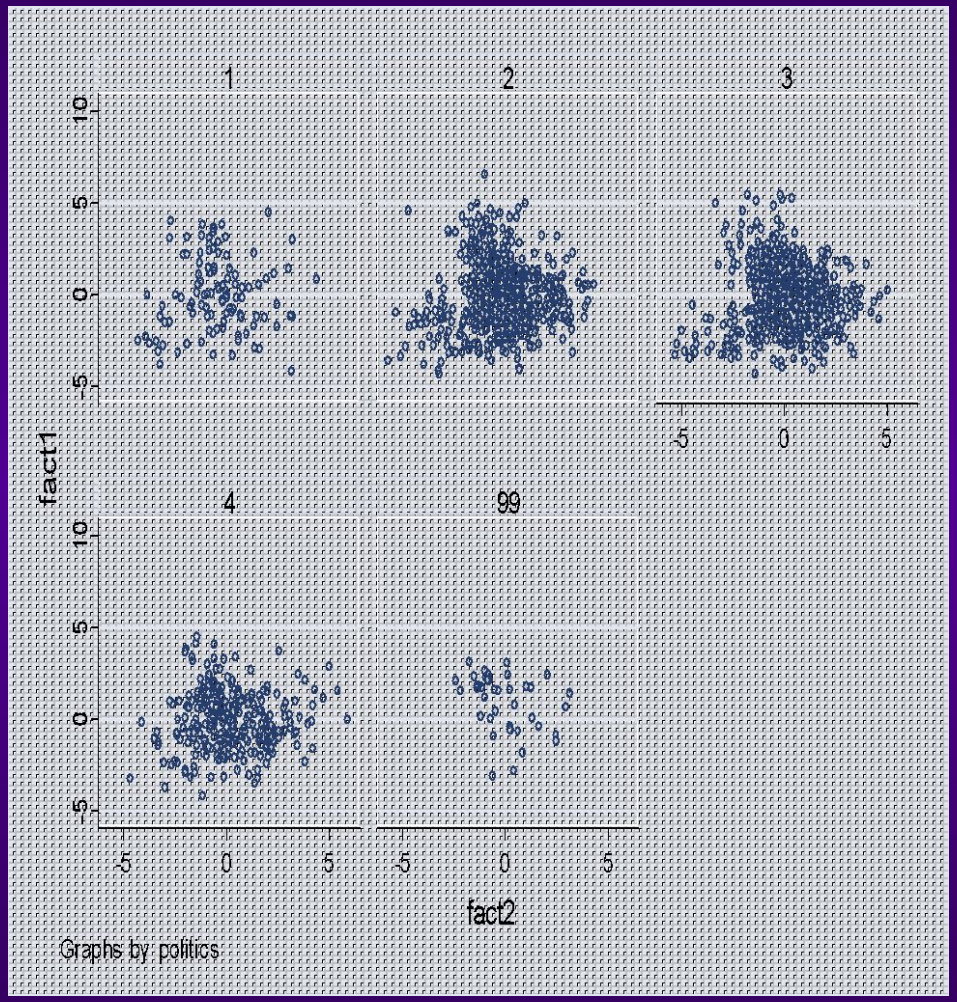


Табл.6. Заинтересованность политикой

Рис.6. 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

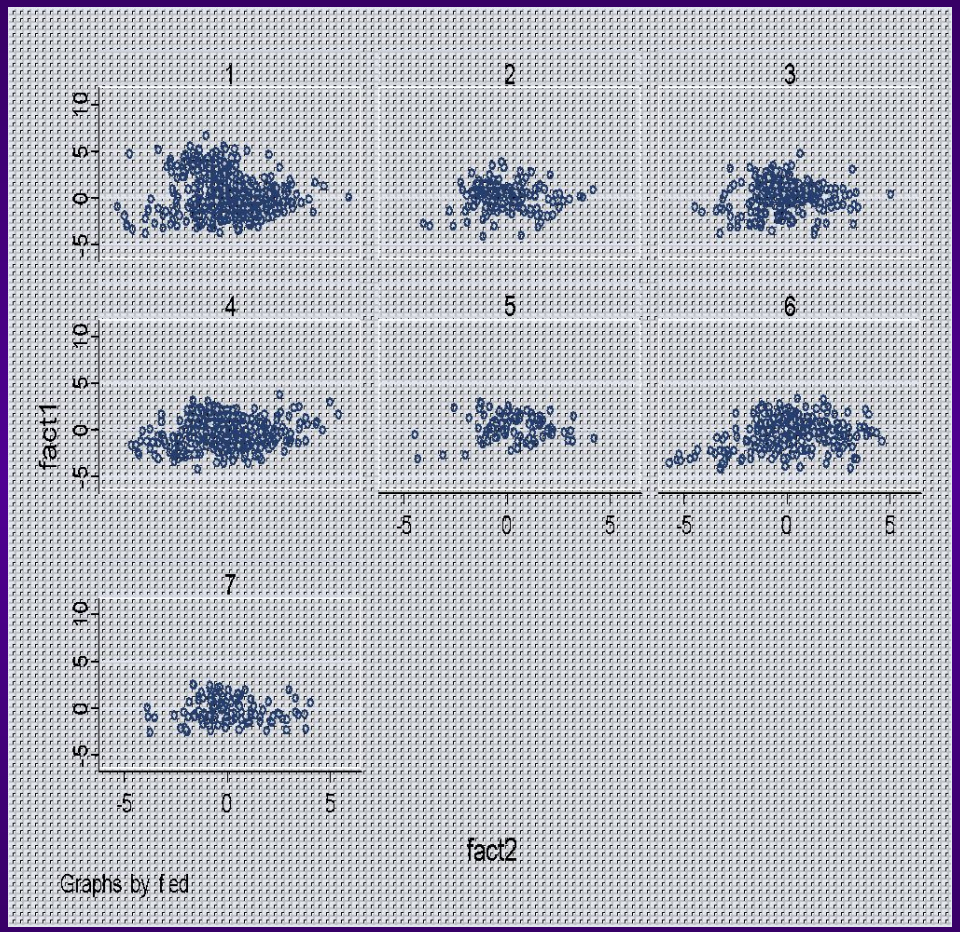


Рис.7. Федеральный округ

6. Регионы

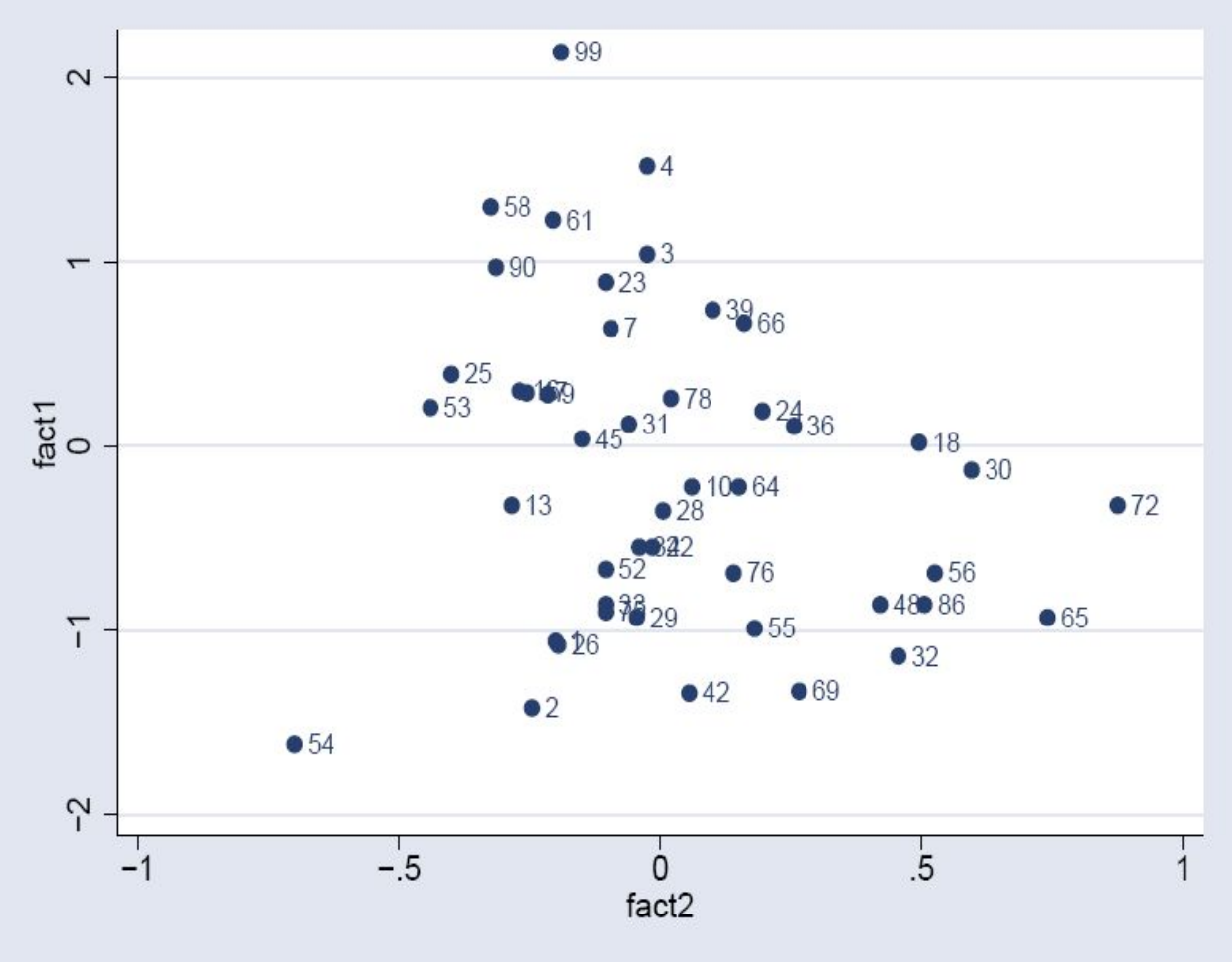


Рис.8. Регионы

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.

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

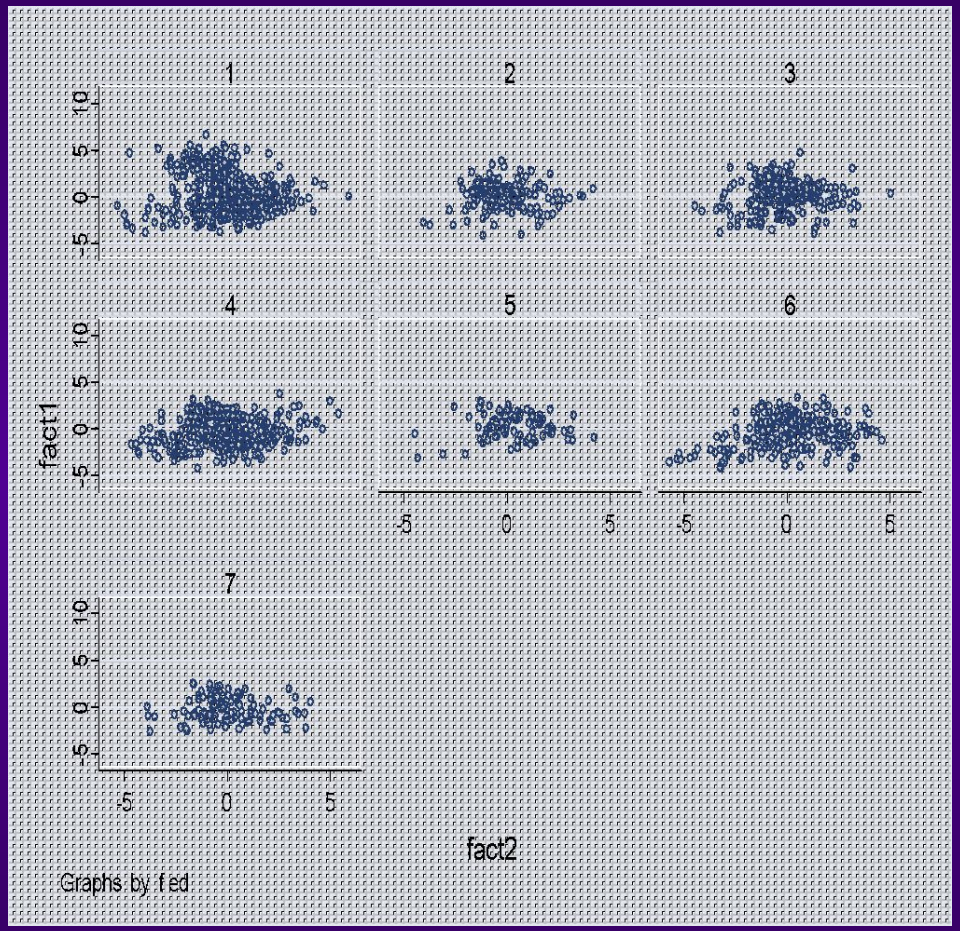


Рис.9. Размер населенного пункта

8. Москва

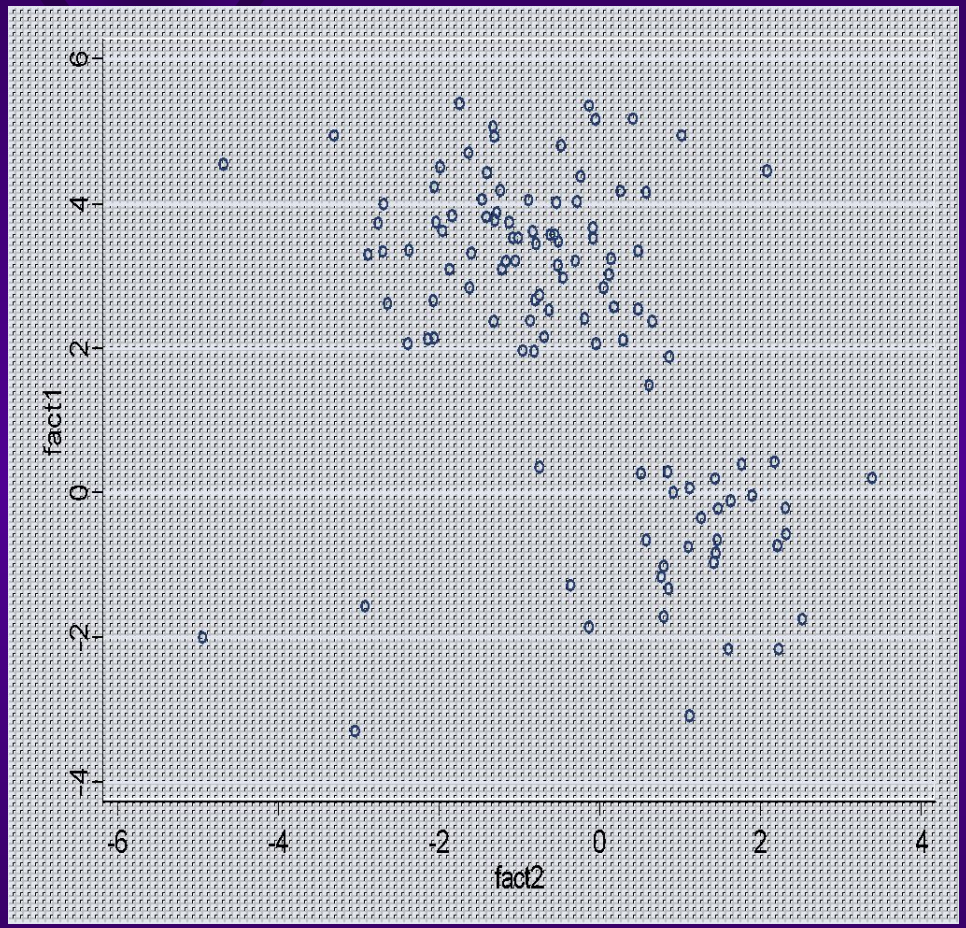


Рис.10. Москва. 2 кластера: недовольные «экономически продвинутые» и довольные «непродвинутые»

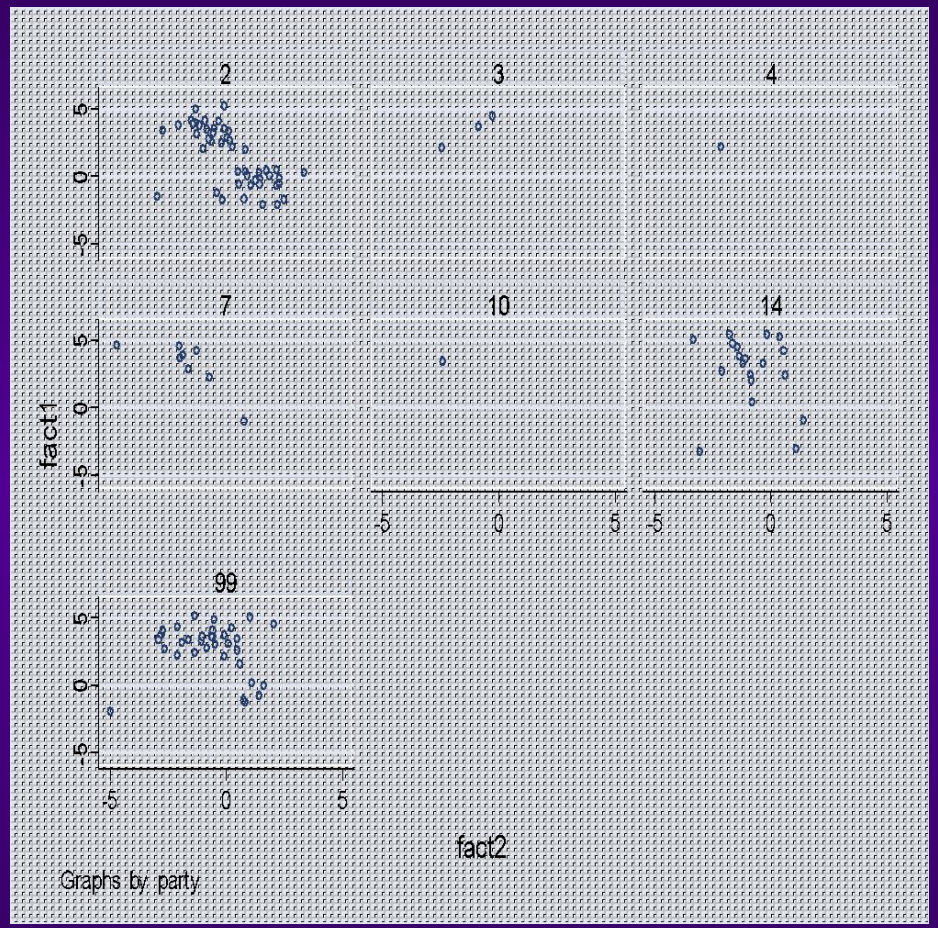


Рис.11. Москва. Партии 2 – ЕР, 3 – КПРФ, 4 – ЛДПР, 7 – СР, 10 – Яблоко, 14 – не голосов.

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

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.9. Доверие к Путину **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

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

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 и доход, недоверие, молодежь, город.

Табл.13. Процент голосующих за партии (факт / наличие «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. Влияние изменения экономического положения за последние годы.

Спасибо за внимание!

alexander.filatov@gmail.com
http://vk.com/alexander.filatov, http://vk.com/baikalreadings