

New Methods for Assessing the Quality of Survey Data

Jörg Blasius

University of Bonn, Germany

Research Methods Festival
Session: Survey Data Quality

Oxford, June 2-5, 2012

There are several ways of task simplification (Blasius and Thiessen, *Assessing the Quality of Survey Data*, 2012, Sage, London):

First, the respondents can simplify their life, which results in, for example, response styles such as acquiescence, disacquiescence, extreme response styles, midpoint responding, ...

Second, the (employees of an) institute can simplify their life, for example by using “copy and paste”-procedures.

Third, the interviewers can simplify their life, for example by faking or partly faking the interviews.

First Example: Detecting Procedural Deficiencies

Simple response structures and duplicates in the *World Value Survey* 2005-2008:

Idea: Applying *Principal Component Analysis* (PCA) to detect simple response structures in 10-point scales on ten items on attitudes towards the democracy in the country. In theory, there are 10^{10} combinations of values but first, the responses are not independent from each other and second, very simple response structures may occur, for example:

10 – 10 – 10 – 10 – 10 – 1 – 1 – 1 – 1 – 1 or

5 – 5 – 5 – 5 – 5 – 5 – 5 – 5 – 5 – 5

Data are available for more than 40 countries, each countries has between 1.000 and 2.500 cases.

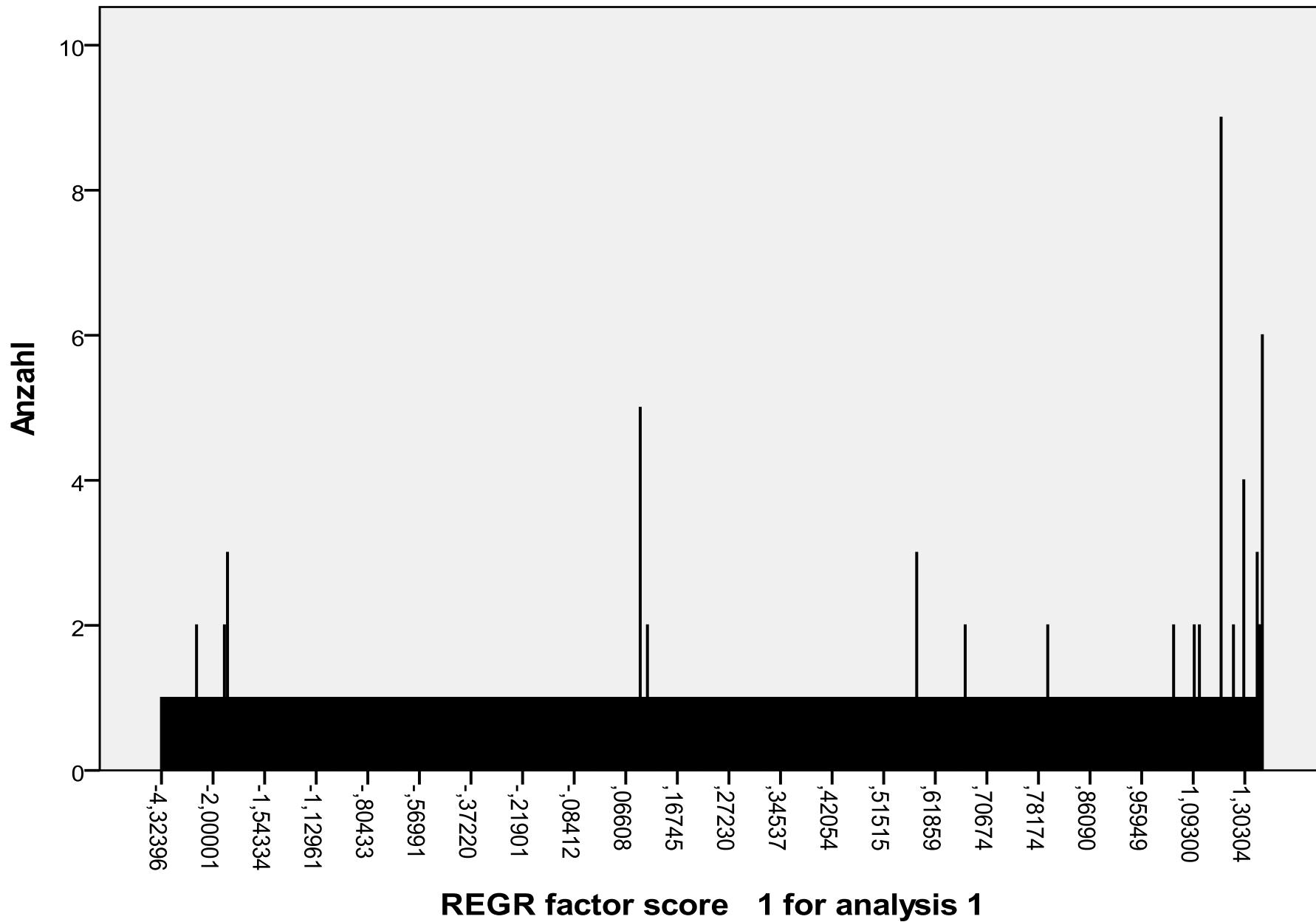
- Governments tax the rich and subsidize the poor.
- Religious authorities interpret the laws.
- People choose their leaders in free elections.
- People receive state aid for unemployment.
- The army takes over when government is incompetent.
- Civil rights protect people's liberty against opposition.
- The economy is prospering.
- Criminals are severely punished.
- People can change the laws in referendums.
- Women have the same rights as men.

All items running from “not at all an essential characteristic of democracy” (1) to “an essential characteristic of democracy” (10)

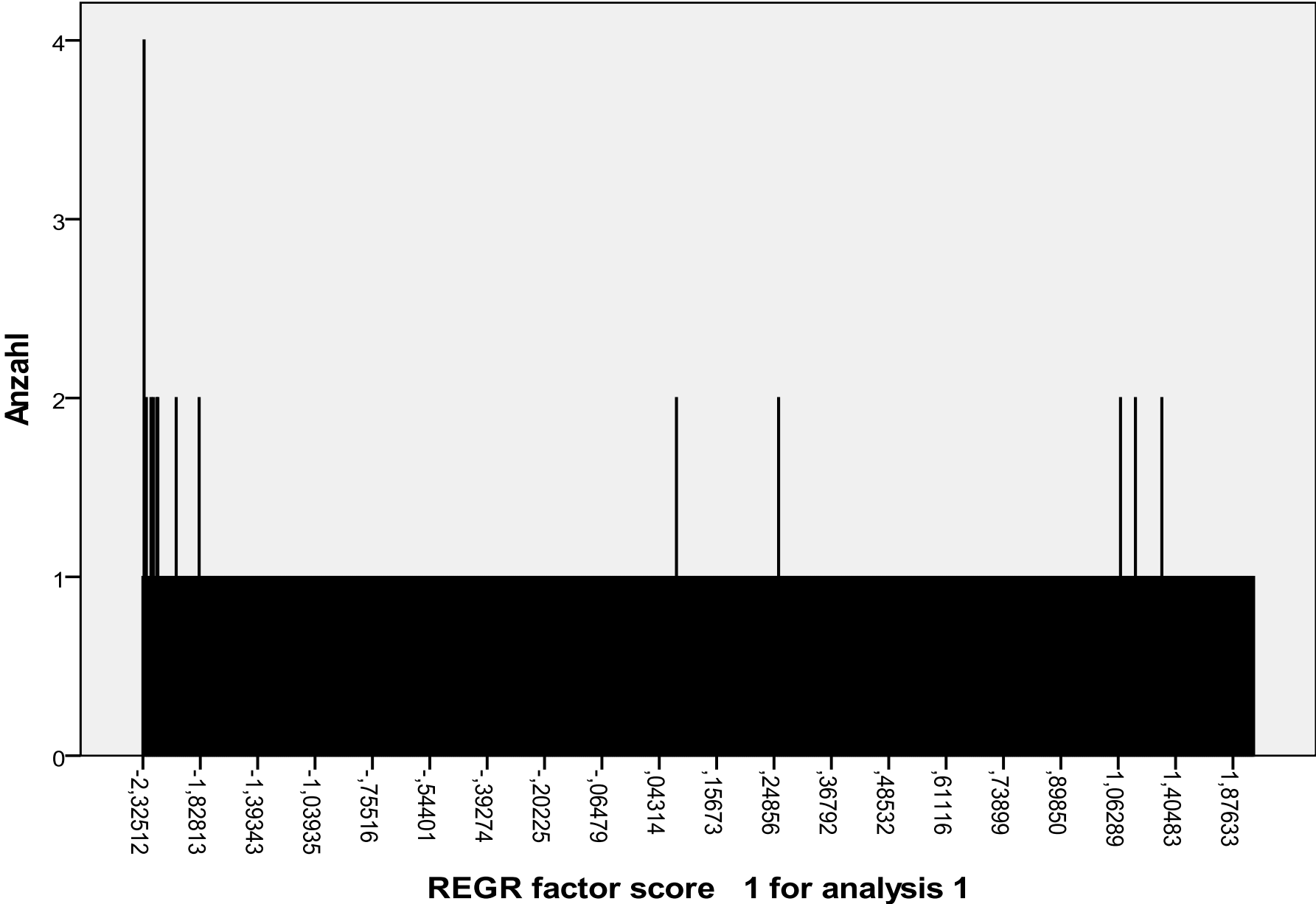
Times each response option was used for the democracy items, in % (+ MD)

Country	1	2	3	4	5	6	7	8	9	10
Australia	4.6	2.0	3.2	3.1	9.7	6.2	8.8	11.7	10.0	38.1
Canada	2.6	1.9	2.8	3.3	8.0	6.7	10.4	17.4	13.2	29.1
China	2.2	1.1	0.9	1.0	2.7	3.2	4.1	10.0	16.8	36.0
Ethiopia	1.7	0.6	0.9	0.8	3.4	4.1	10.3	15.4	23.2	35.2
Finland	1.7	1.7	2.8	3.3	7.5	7.4	12.9	20.0	16.9	24.1
Germany	1.9	1.0	1.7	2.4	5.7	5.9	8.2	12.5	10.5	47.8
Ghana	10.2	3.3	3.7	2.9	4.6	6.4	7.9	12.2	15.2	31.2
India	8.8	0.0	0.0	0.0	16.1	0.0	0.0	0.0	0.0	53.5
Mexico	14.4	3.8	3.7	3.4	8.2	5.3	6.5	10.3	8.5	28.2
Norway	5.5	3.4	4.0	3.0	7.8	5.0	9.4	15.4	13.5	31.2
South Africa	3.7	2.8	2.5	3.5	8.1	6.7	8.7	11.6	13.3	35.4
Spain	2.3	1.0	1.5	2.6	8.1	6.6	12.1	11.9	10.3	38.3
Turkey	2.3	1.0	1.3	2.0	5.6	5.4	10.1	16.5	8.7	44.1
USA	5.3	2.4	3.4	3.8	14.5	8.1	9.3	10.0	9.1	30.0
Zambia	5.6	3.4	4.4	4.5	10.7	8.1	6.5	7.5	8.9	36.4

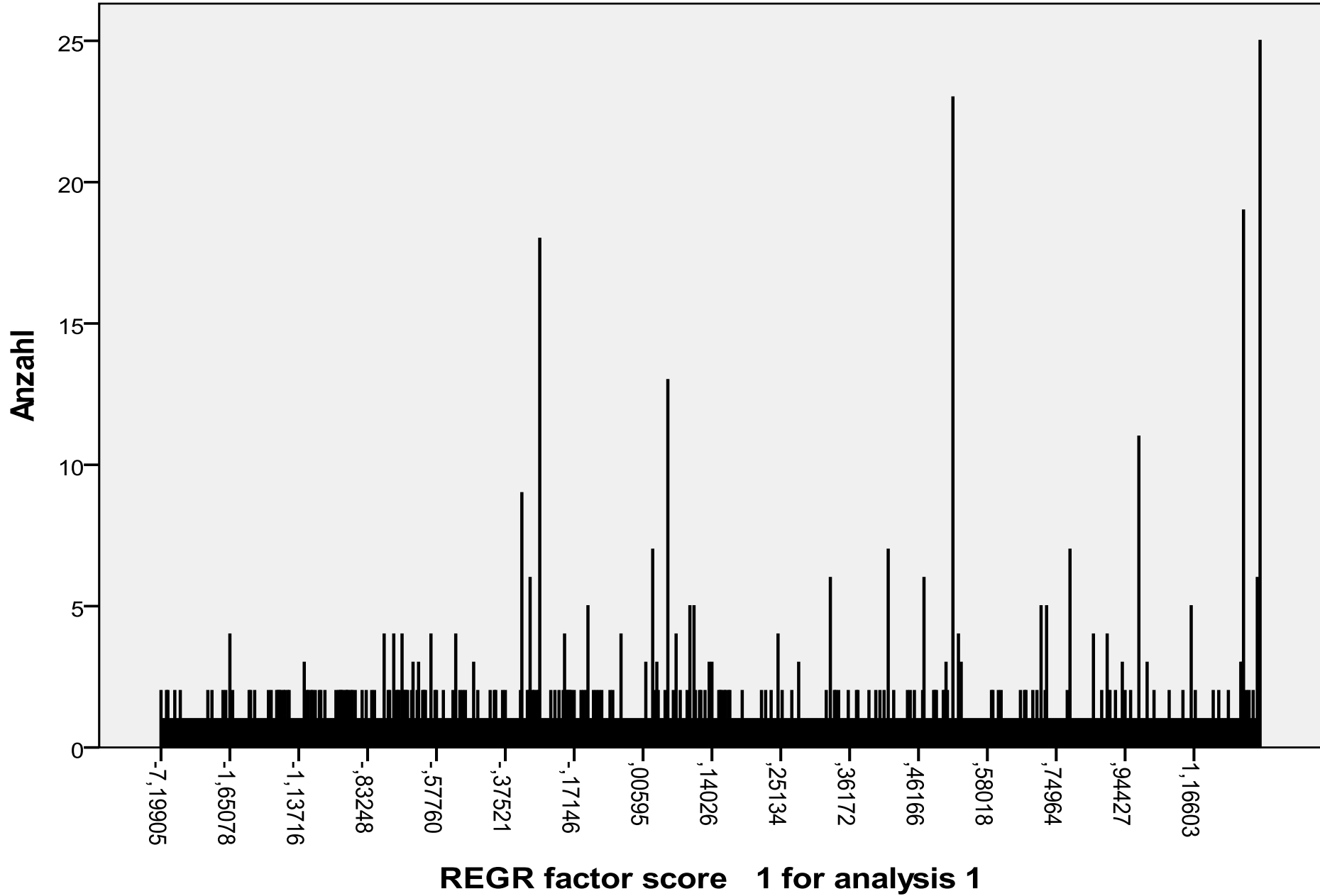
Country/Region: USA



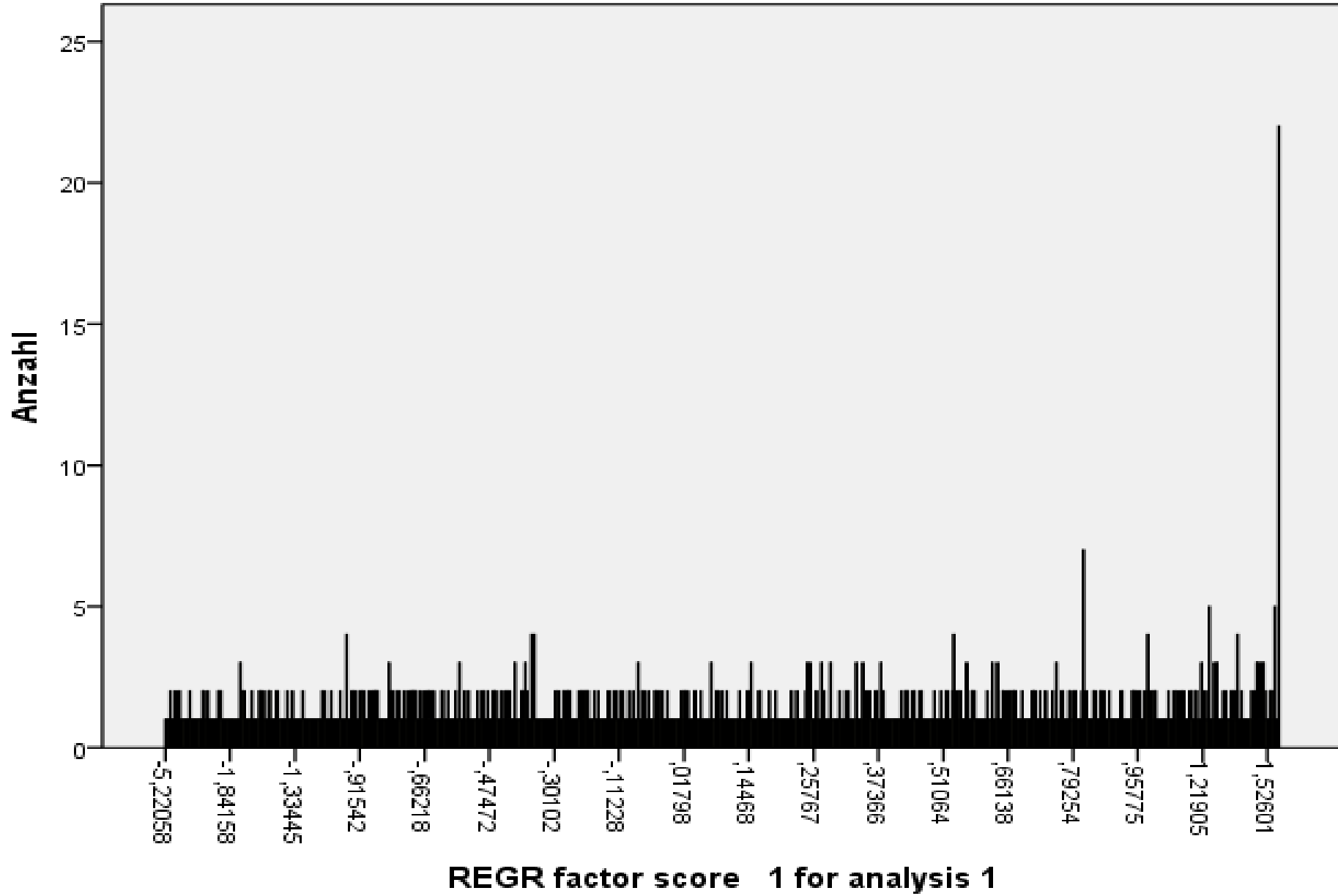
Country/Region: Norway



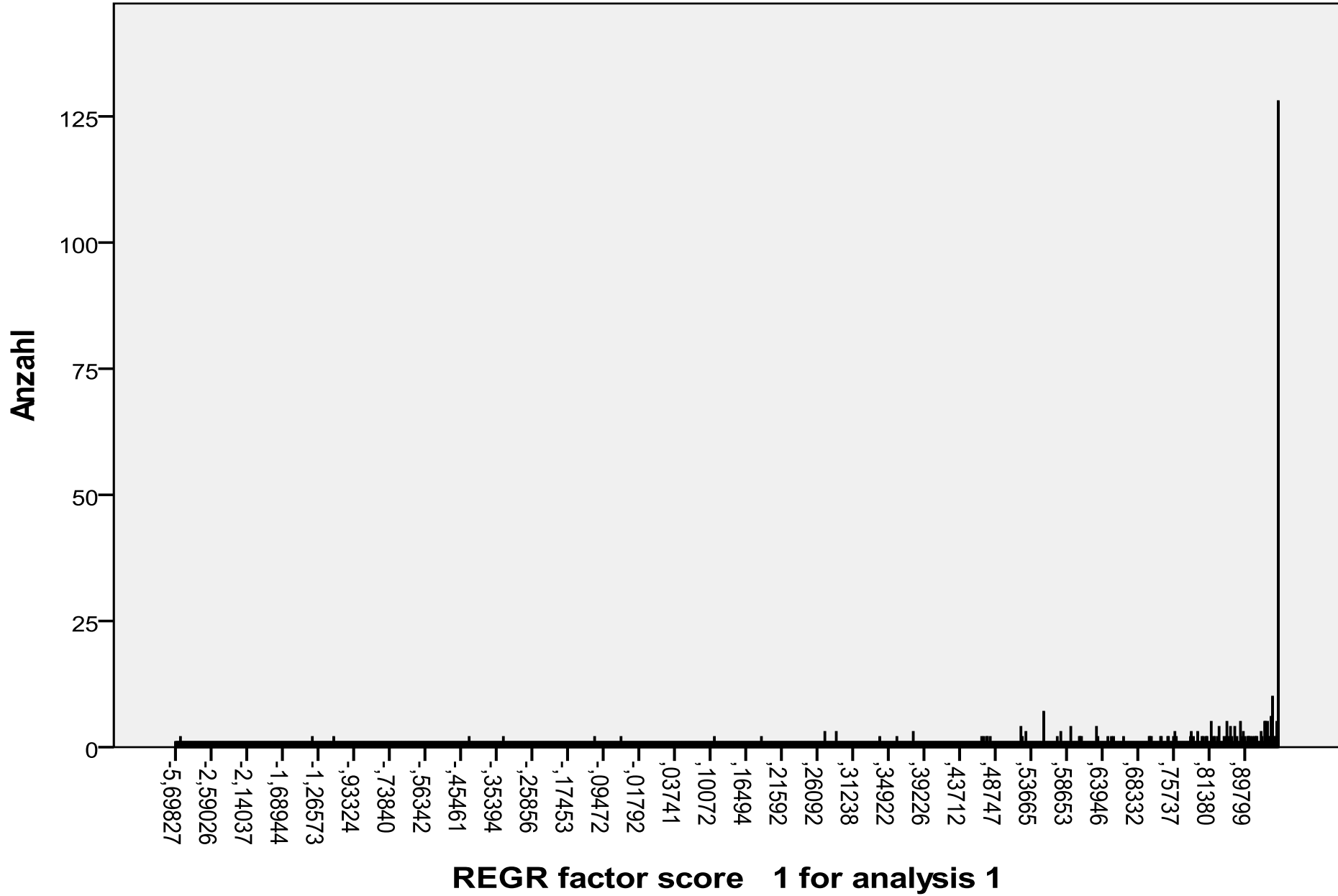
Country/Region: Ethiopia



Country/Region: S Korea



Country/Region: Germany



Response patterns by country

Pattern 1 (Pat. 1): 10-1-10-10-1-10-10-10-10-10

Pattern 2 (Pat. 2): 10-10-10-10-1-10-10-10-10-10

Country	N	% MD	Pat. 1	Pat. 2	% cases in max.co-occur.	N of dup. combin.	% dupl. cases	Duplicates - combination ratio
Andorra	1003	4.2	51	1	5.3	53	19.3	3.5
Argentina	1002	31.6	16	23	3.4	41	15.3	2.6
Australia	1421	6.8	9	4	0.7	29	3.5	1.6
Brazil	1500	15.7	6	3	0.9	27	5.1	2.4
Bulgaria	1001	33.3	11	1	2.7	17	7.9	3.1
Burkina Faso	1534	27.6	12	8	1.5	47	11.7	2.8
Canada	2164	20.0	5	1	0.3	14	1.3	1.6
Chile	1000	23.1	13	8	1.7	27	6.5	1.9
China	2015	62.3	17	1	3.0	20	7.5	2.9
Cyprus	1050	1.5	36	13	3.5	53	14.8	2.9
Egypt	3051	6.3	2	33	1.2	151	16.2	3.1
Ethiopia	1500	18.0	19	25	2.0	220	33.8	1.9
Finland	1014	6.5	7	0	0.7	4	1.1	2.5
Georgia	1500	35.2	15	4	2.2	59	16.4	2.7

Germany	2064	10.9	128	2	7.0	95	15.6	3.0
Ghana	1534	12.8	0	0	1.6	39	7.1	2.4
India	2001	48.6	49	2	4.8	132	38.4	3.0
Indonesia	2050	24.9	3	5	1.5	104	14.7	2.2
Japan	1096	35.9	15	0	2.1	16	4.4	1.9
Jordan	1200	20.8	2	9	1.1	43	9.4	2.1
Malaysia	1201	0.7	0	1	0.5	50	6.3	1.5
Mali	1534	29.5	4	6	3.7	48	15.5	3.5
Mexico	1560	16.3	5	1	1.4	28	5.7	2.6
Moldova	1046	16.0	8	3	0.9	61	9.0	1.3
Morocco	1200	32.8	3	0	0.6	17	4.0	1.9
Norway	1025	4.0	2	1	0.4	13	1.5	1.2
Peru	1500	18.8	8	6	0.7	19	4.3	2.7
Poland	1000	22.3	14	4	1.8	19	5.9	2.4
Romania	1776	24.5	27	10	2.0	72	14.4	2.7
S Africa	2988	19.8	5	17	0.7	38	4.1	2.6
S Korea	1200	0.2	22	3	1.8	254	27.0	1.3
Serbia	1220	23.6	13	5	1.4	47	8.2	1.6
Slovenia	1037	23.6	13	3	1.6	16	4.4	2.2
Spain	1200	18.4	48	4	4.9	39	14.0	3.5
Sweden	1003	6.9	8	0	1.0	29	6.5	2.1
Switzerland	1241	12.7	17	0	1.6	36	7.0	2.1
Taiwan	1227	3.1	17	1	1.4	30	5.4	2.1

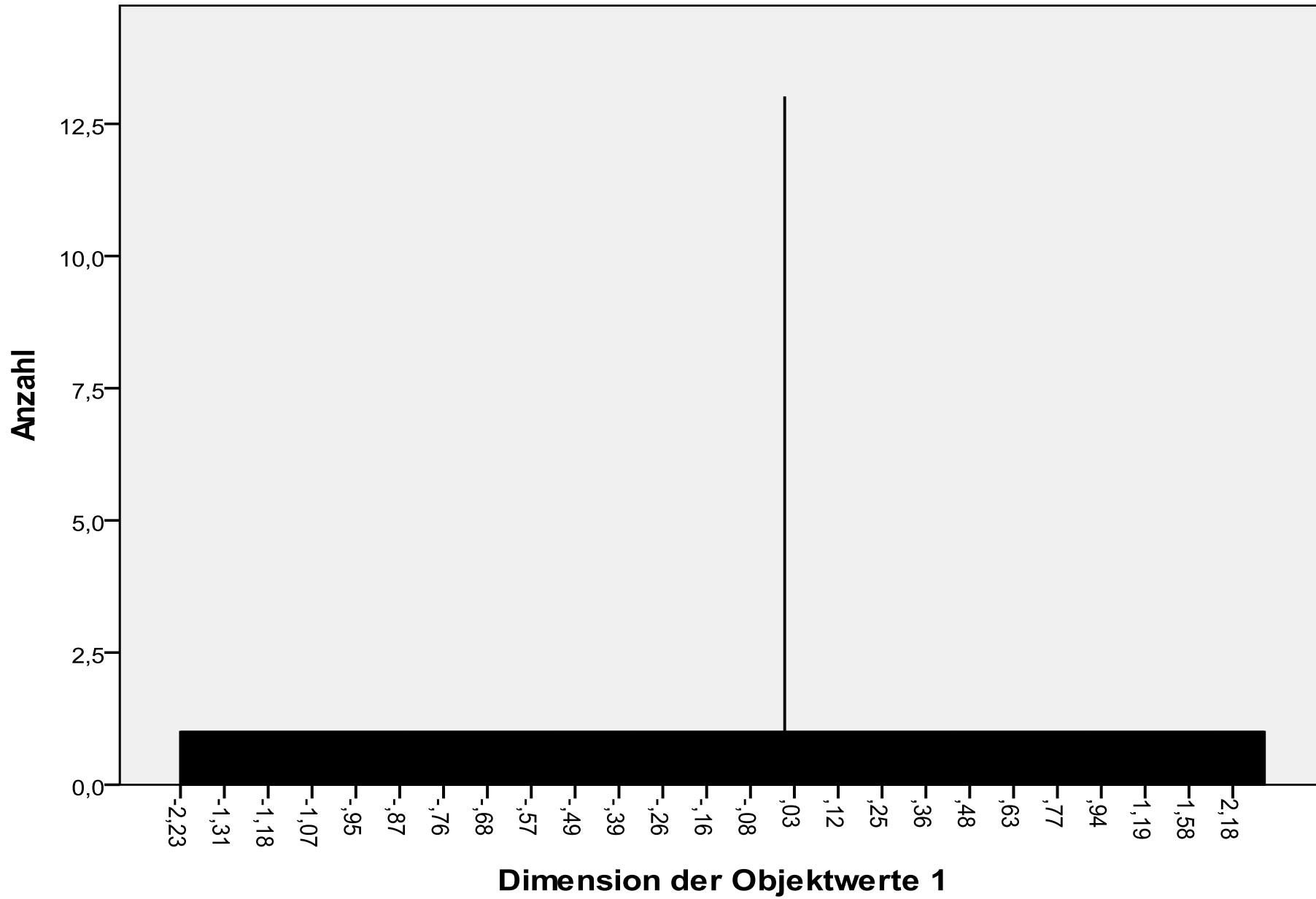
Thailand	1534	2.3	3	0	0.7	67	6.7	1.5
Trinidad	1002	8.7	9	2	1.4	19	5.2	2.5
Turkey	1346	10.9	20	6	1.7	45	11.5	3.1
Ukraine	1000	32.3	7	6	1.0	15	5.0	2.3
Uruguay	1000	10.0	21	9	2.3	18	6.2	3.1
USA	1249	6.2	6	0	0.8	19	3.2	2.0
Vietnam	1495	23.3	3	51	4.5	49	13.5	3.2
Zambia	1500	16.0	3	7	0.6	65	8.3	1.6

Second Example: Data Duplication (Copy and Paste)

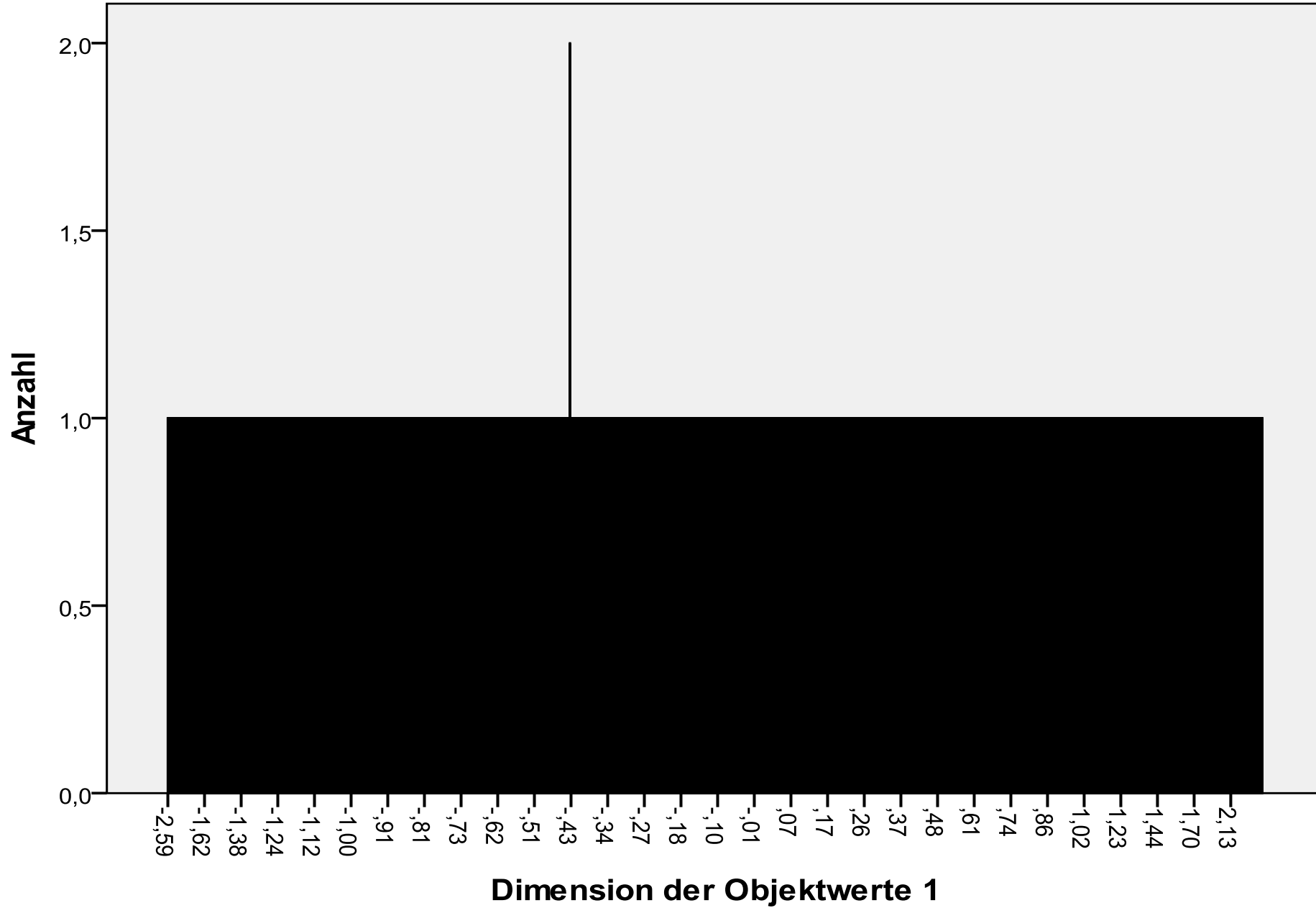
Using a large number of variables from different areas will not provide with any meaningful scale, but if “respondents” gave the same answers they receive the same scale values. We use 36 variables, including eight four-point variables on *gender roles*, one 10-point variable on *satisfaction with the financial situation*, two ranking scales of four choices on *national goals*, six ranking scales of four choices each on *materialism and post-materialism*, 10 six-point *self-description* variables, and four 10-point variables on *technology* (v60-v95). There should be no two respondents receiving the same scale values.

For including the missing values as valid categories we use *Multiple Correspondence Analysis* (MCA).

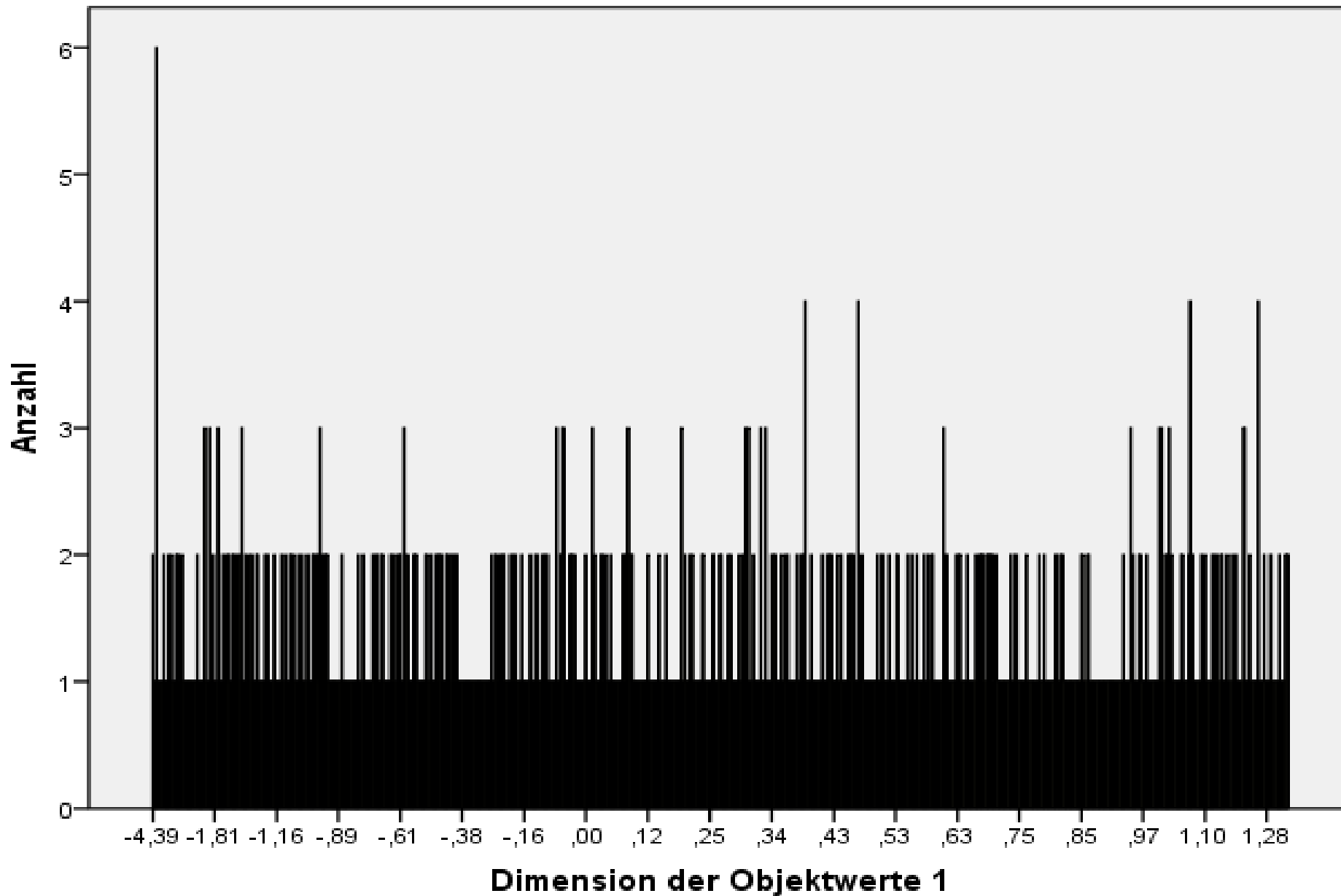
Country/Region: USA



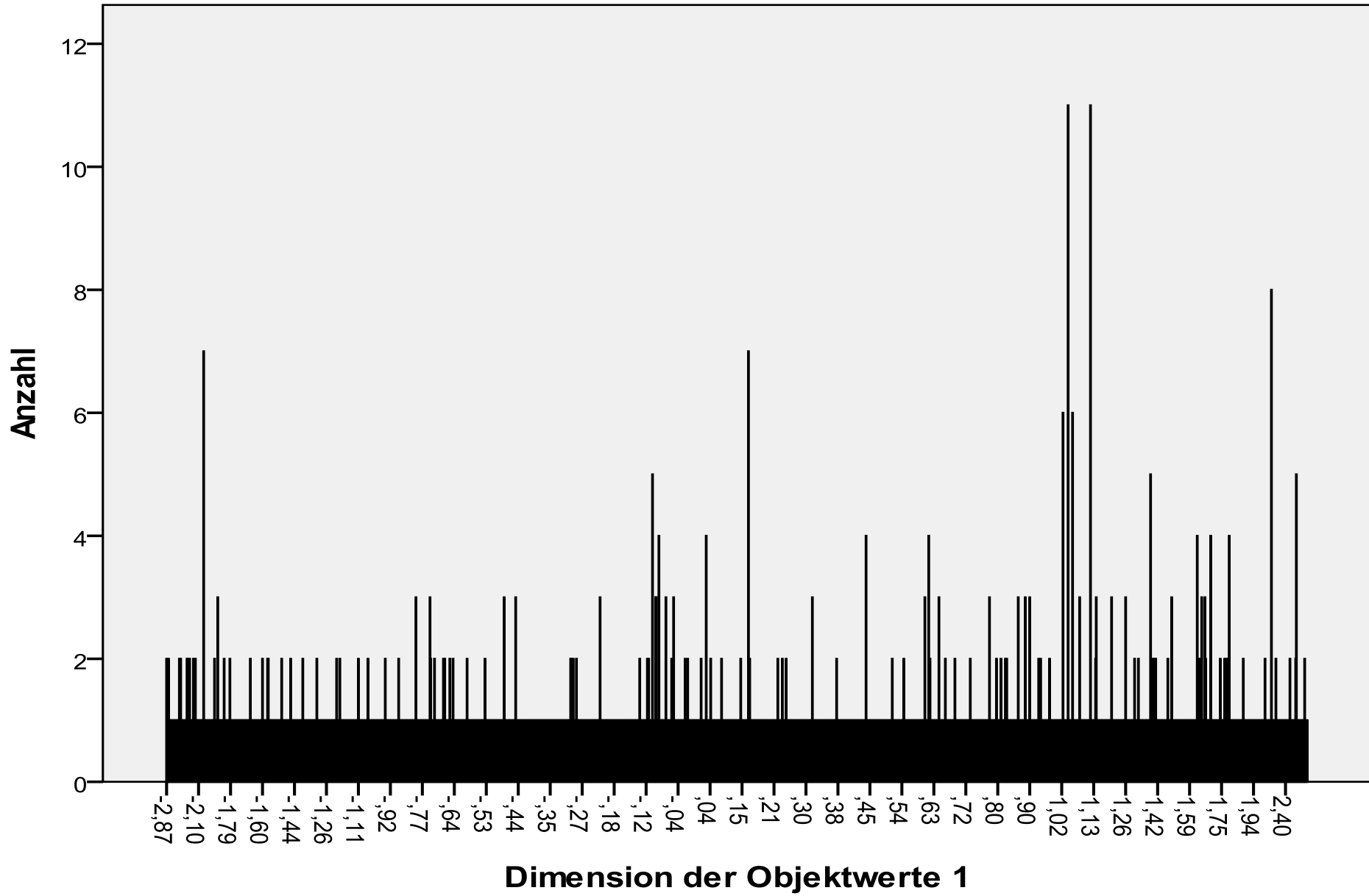
Country/Region: S Africa



Country/Region: S Korea



Country/Region: India



Calculation of duplicates for selected countries

k	South Korea		Indonesia		Thailand		India		Moldova		Ethiopia	
	f1	f2	f1	f2	f1	f2	f1	f2	f1	f2	f1	f2
1	671	671	1922	1922	1454	1454	1659	1659	998	998	835	835
2	219	438	26	52	32	64	85	170	24	48	254	508
3	23	69	7	21	1	3	24	73			16	48
4	4	16	3	12			7	28			7	28
5					1	5	3	15			4	20
6	1	6					2	12			1	6
7							2	14			2	14
8			1	8	1	8	1	8				
10											2	20
11							2	22				
21											1	21
N	918	1200	1959	2015	1489	1534	1785	2001	1022	1046	1122	1500
Fakes		282		56		45		216		24		378

Key: k = Frequency of occurrence; f1 = number of instances; f2 = number of cases represented. The number of duplicates or fakes is obtained by subtracting the N in f1 from the N in f2.

Third Example: Response Quality in Different Countries

The International Social Survey Program (ISSP)

- One of the best international surveys
- More than 40 countries from all continents participate
- Sample size should exceed 1,000 interviews in each country
- More than 1,000 journal articles using the ISSP data,
- And, more than 500 book chapters and books, conference papers, student projects and all other kinds of reports

1994 data, where the focus was on “Family and Changing Gender Roles”, in this year 22 countries participated. The data set contains 11 variables with five response categories each (from strongly agree to strongly disagree).

Items measuring support for single/dual earner family structure

- A A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.
- B A pre-school child is likely to suffer if his or her mother works
- C All in all, family life suffers when the woman has a full-time job
- D A job is all right, but what most women really want is a home and children.
- E Being a house wife is just as fulfilling as working for pay.
- F Having a job is the best way for a woman to be an independent person.
- G Most women have to work these days to support their families.
- H Both the man and woman should contribute to the household income.
- I A man's job is to earn money; a woman's job is to look after the home and family
- J It is not good if the man stays at home and cares for and the woman goes out to work.
- K Family life often suffers because men concentrate too much on their work.

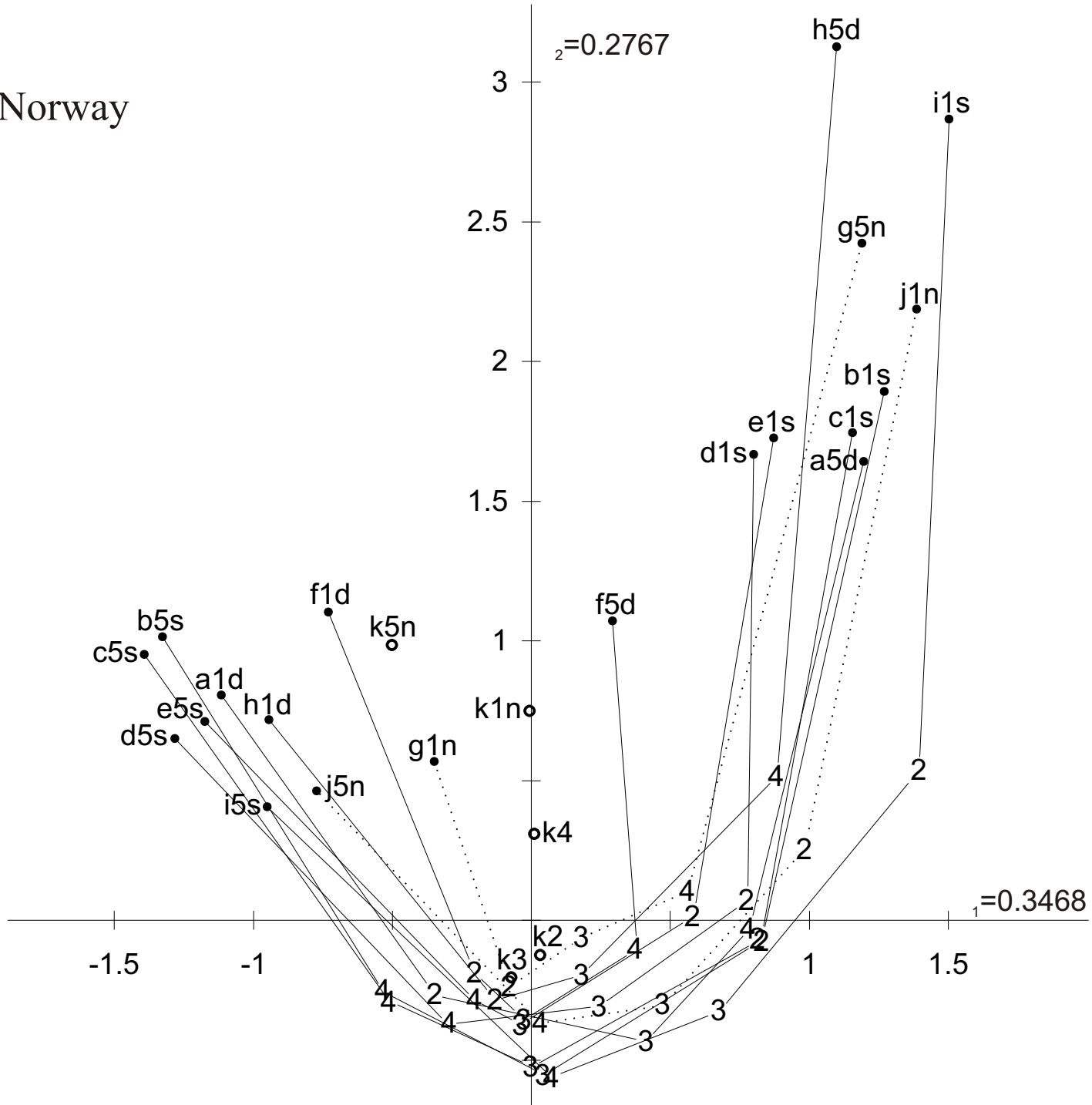
Constructing a two-dimensional Map by Means of Multiple Correspondence Analysis

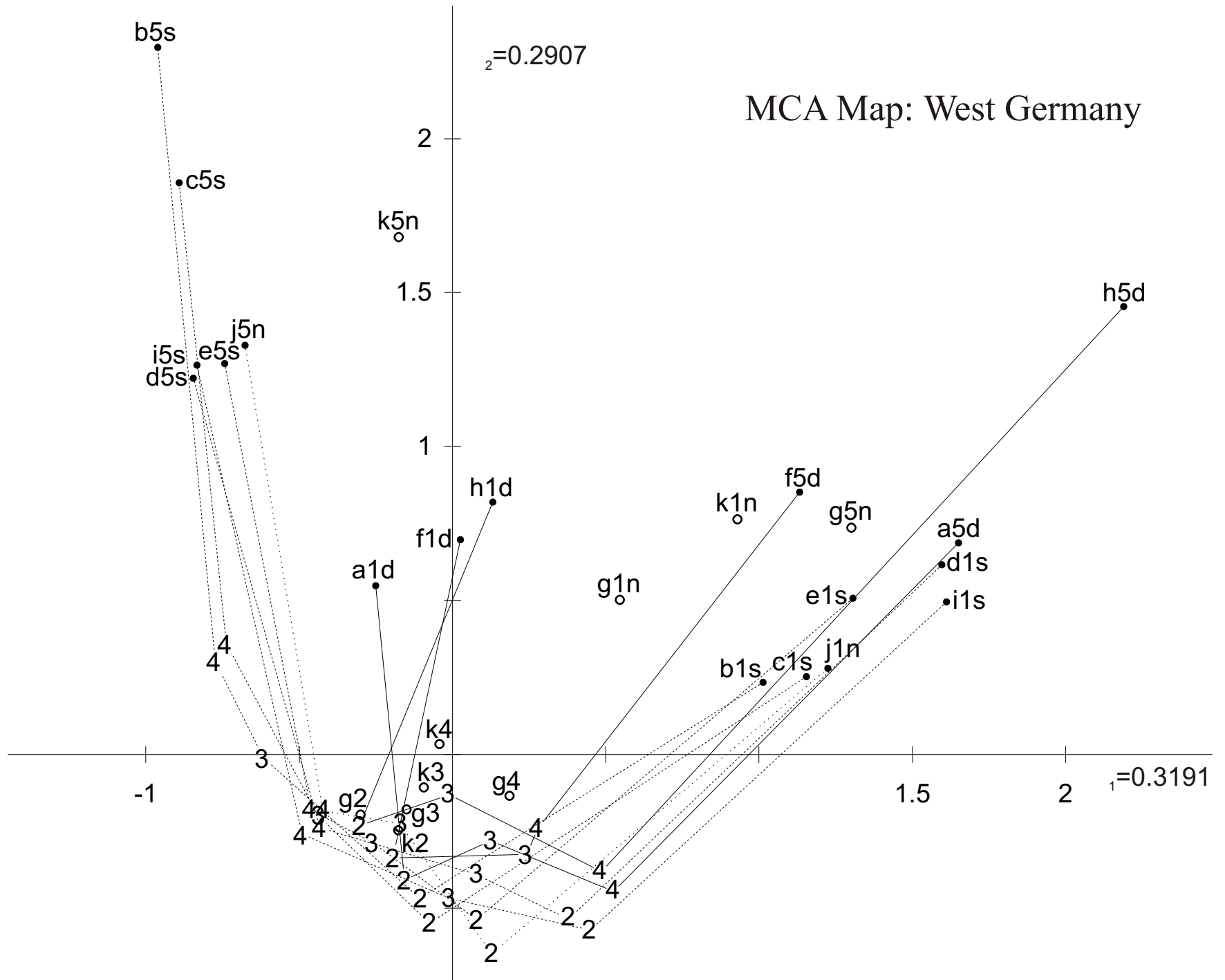
- “The model should follow the data, not the inverse” (Jean-Paul Benzécri).
- Except that non-negative entries are not allowed (which does not exist in survey data), there are no constraints on the data.
- Best method to see different kinds of methodologically-induced variation, for example, response sets; as well as to distinguish between methodologically-induced and substantive variation.
- Similarities between variable categories (and as well between respondents) are reflected by short (Euclidian) distances, dissimilarities by large distances.
- When using ordinal data, the second dimension often mirrors a method-induced horseshoe (arch or Guttman effect).

Possible Outcomes in MCA

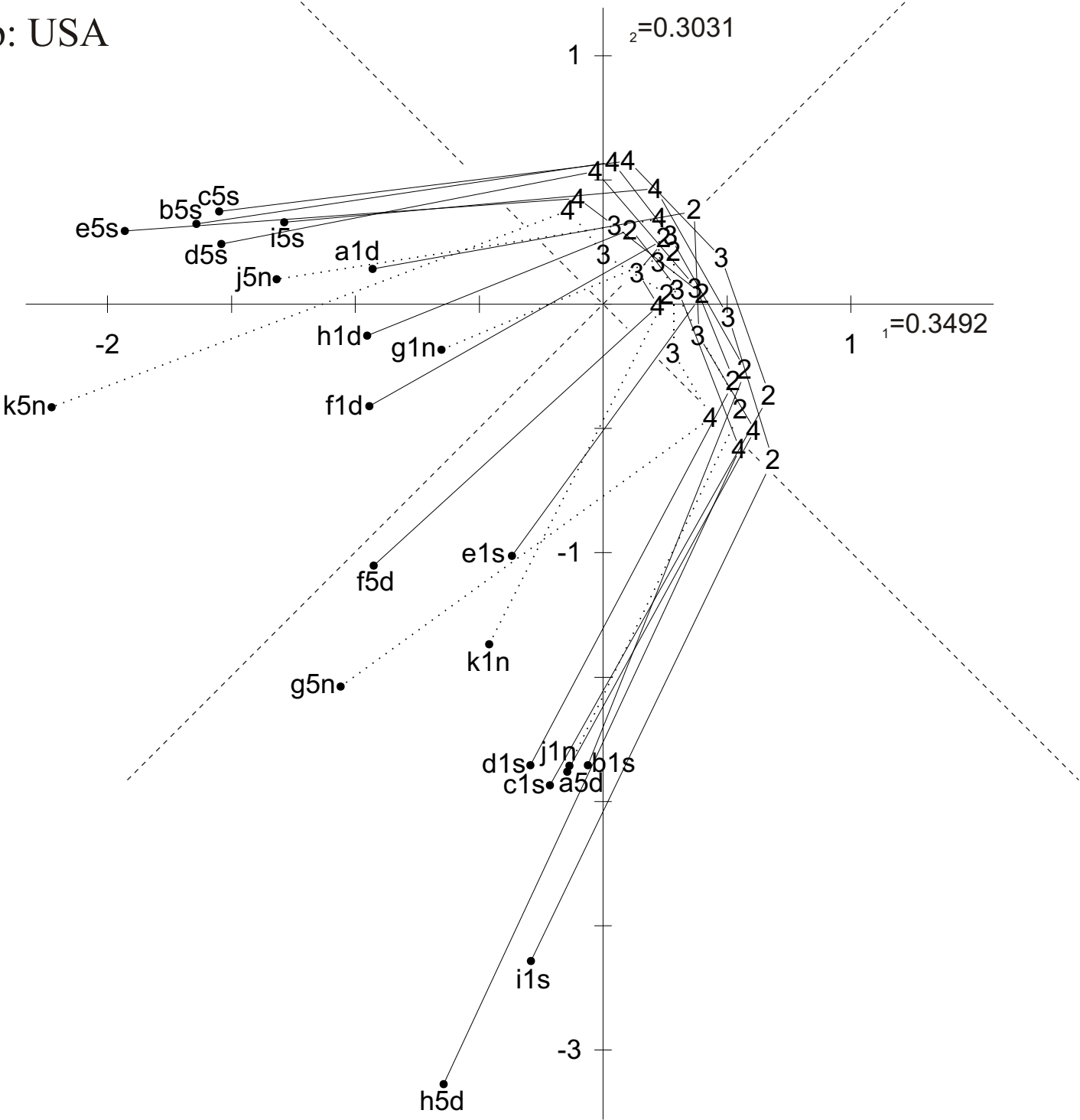
- If the quality of data is high, the first (two) dimension(s) should capture mainly substantive variation due to gender roles, with the second (third) dimension reflecting the horseshoe.
- Items that neither measure support for single nor for dual earner households might be manifested in higher dimensions.
- The items associated with the first (two) dimension(s) should retain their ordinality in this (these) dimension(s).
- If the question is formulated in reversed order (by country), the responses to the reversed-formulated item will be reversed as well.
- The horseshoe might also appear on the first dimension (large amount of non-substantive variation) or between dimensions 1 and 2 (two-dimensional solution, data might be on high quality).

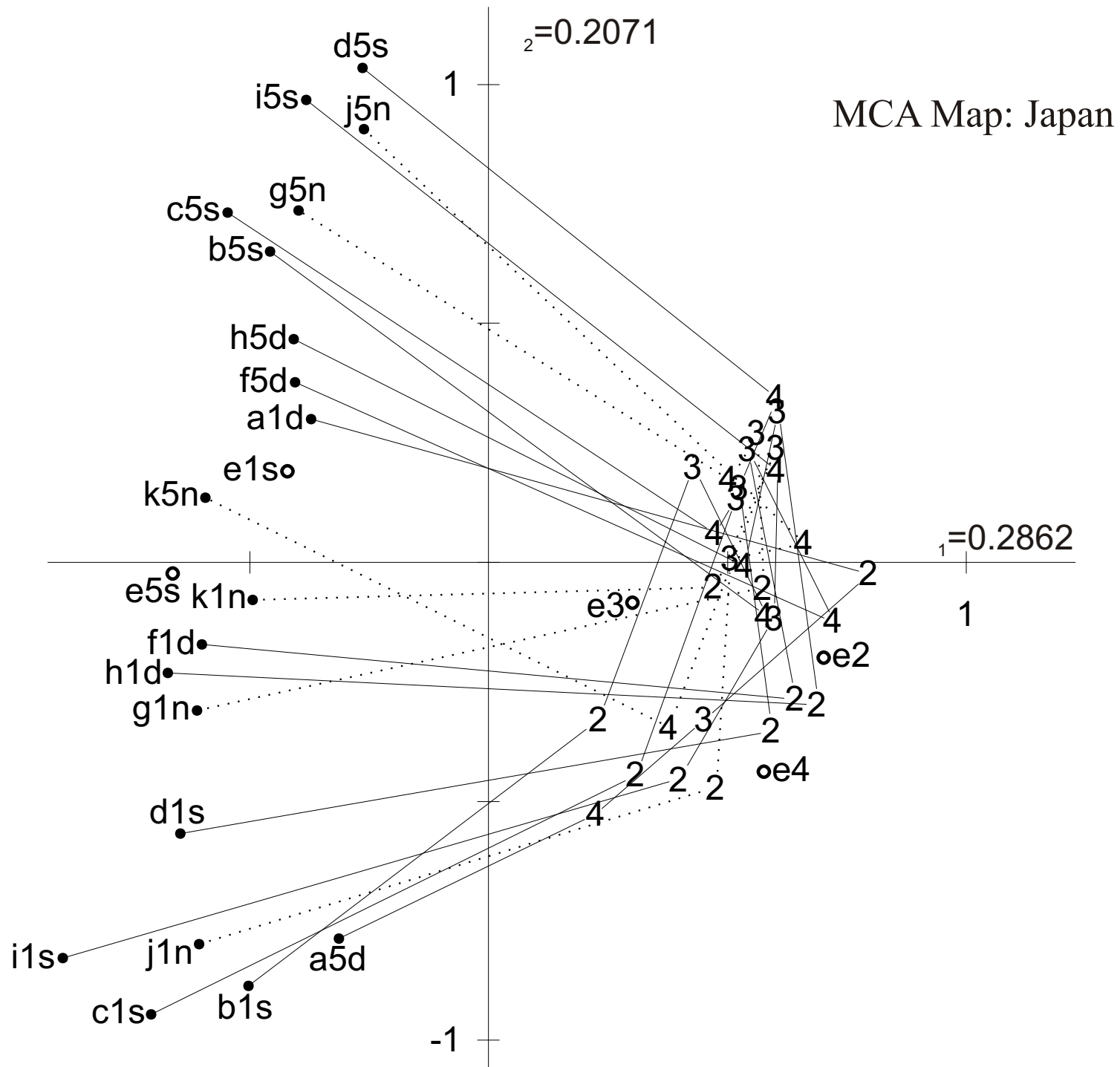
MCA Map: Norway



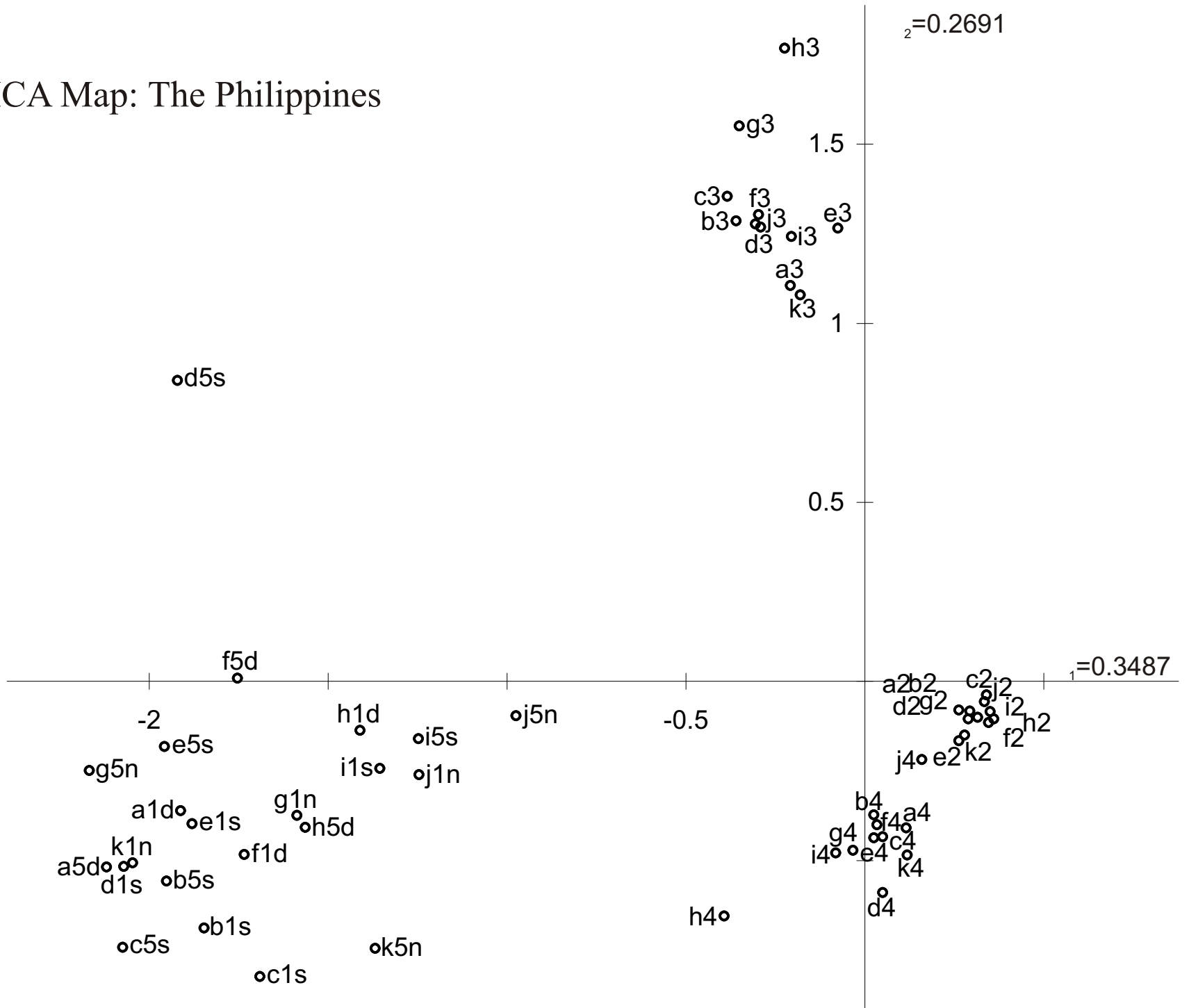


MCA Map: USA





MCA Map: The Philippines



Fourth Example: Response Quality and Moderator Effects

The *1984 Canadian National Election Study* (CNES) contains items on *political efficacy and trust* that lend themselves well to exploring the connections between political interest and education on the one hand, and data quality on the other. Balch (1974) first introduced these items, spawning voluminous research based on variations of his items, including national election studies, several ISSP surveys, and other cross-national studies. Regardless of the particular operationalization of political efficacy and trust, positive relationships with both education and political interest are generally reported. However, unequal data quality might produce artifactual relations that masqueraded as substantive ones.

Canadian Nationwide Election Study 1984: “Political Trust and Efficacy Data” (N=3,377)

Item	SA	AS	NN	DS	SD	NO
a) Generally, those elected to Parliament soon lose touch with the people.	26.6	44.5	3.5	16.1	4.8	4.5
b) I don't think the (Federal) Government cares much about what people like me think.	26.9	32.9	3.8	24.2	9.0	3.2
c) Sometimes, (Federal) Politics and Government seem so complicated that a person like me can't really understand what's going on.	30.8	33.1	2.5	19.1	12.6	1.9
d) People like me don't have any say about what the Government in (Ottawa) does.	33.4	28.3	2.2	20.0	14.0	2.1
e) So many other people vote in (Federal) elections that it does not matter very much whether I vote or not.	7.8	9.9	1.8	16.0	62.8	1.7
f) Many people in the (Federal) Government are dishonest.	10.5	25.1	10.1	24.6	18.2	11.5
g) People in the (Federal) Government waste a lot of the money we pay in taxes.	46.3	33.2	3.9	9.0	3.6	4.1

Item	SA	AS	NN	DS	SD	NO
h) Most of the time we can trust people in the (Federal) Government to do what is right.	10.4	46.0	6.2	23.5	9.7	4.2
i) Most of the people running the (Federal) Government are smart people who usually know what they are doing.	15.9	45.5	5.9	21.0	8.2	3.6

Assumptions on the structure of data

In case of response sets (as a kind of task simplification), the variance in responses for each respondent will decrease while the variation between the respondents will increase. It follows, total inertia will increase.

Further, the correlations (absolute values) between the response categories increase (positive values between the same categories, negative values between the opposite categories) while the structure of responses becomes simpler. This simple structure will be captured by the first MCA/SMCA dimension. It follows: The simpler the structure of responses, the more variance will be explained by the first dimension.

Subset Multiple Correspondence Analysis (SMCA)

SMCA concentrates on just some of the response categories, while excluding others from the solution (Greenacre and Pardo 2006, Greenacre 2007). For example, with SMCA the structure of the subset of NOs can be analyzed separately, or these responses can be excluded from the solution while concentrating only on the substantive responses.

Suppose we have five variables with four categories, ranging from SA to SD. Since the row sums of the indicator matrix are 5, SMCA maintain the equal weighting of all respondents, the row profile values are 0.2 and zero. If we concentrate on SA, respondents with five answers on SA will have five profile values of 0.2 (and a row sum of 1.0), respondents with four answers on SA will have four profile values of 0.2 (and a row sum of 0.8), respondents with two answers on SA will have two profile values of 0.2 (and a row sum of 0.4); in case of omitting the categories they would have four profile values of 0.25 (or two values of 0.5) and a row sum of one.

SMCA, Burt-Table

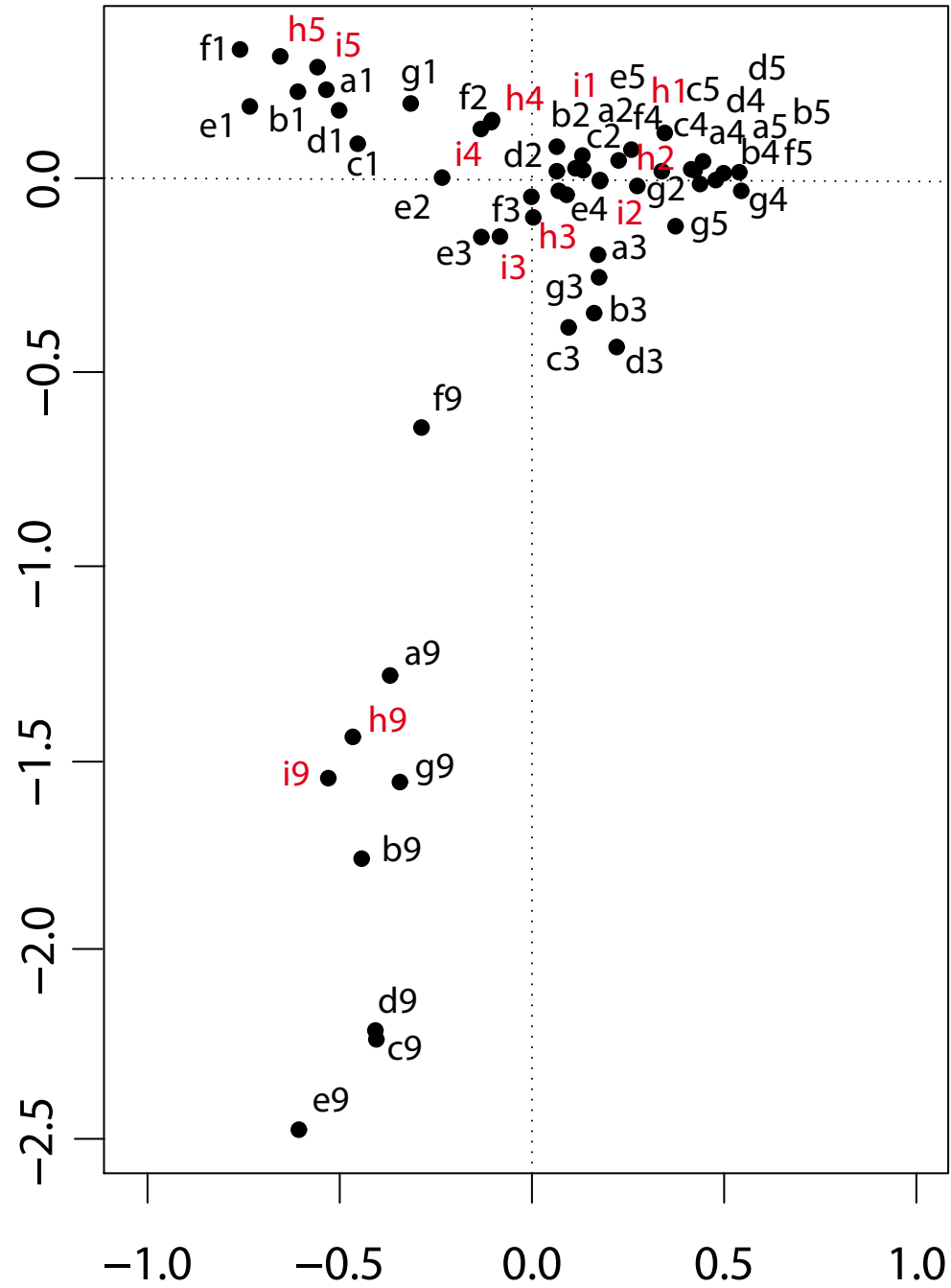
	a1 a2 a4 a5 b1 b2 ... i1 i2 i4 i5	a3 a9 b3 b9 c3 c9 ... i3 i9
a1 a2 a4 a5 b1 b2 ... i4 i5	Subset MCA, Set 1	Interaction, Set 1 × Set 2
a3 a9 b3 b9 ... i3 i9	Interaction, Set 2 × Set 1	Subset MCA, Set 2

Constructing a two-dimensional Map by Means of (Subset) Multiple Correspondence Analysis

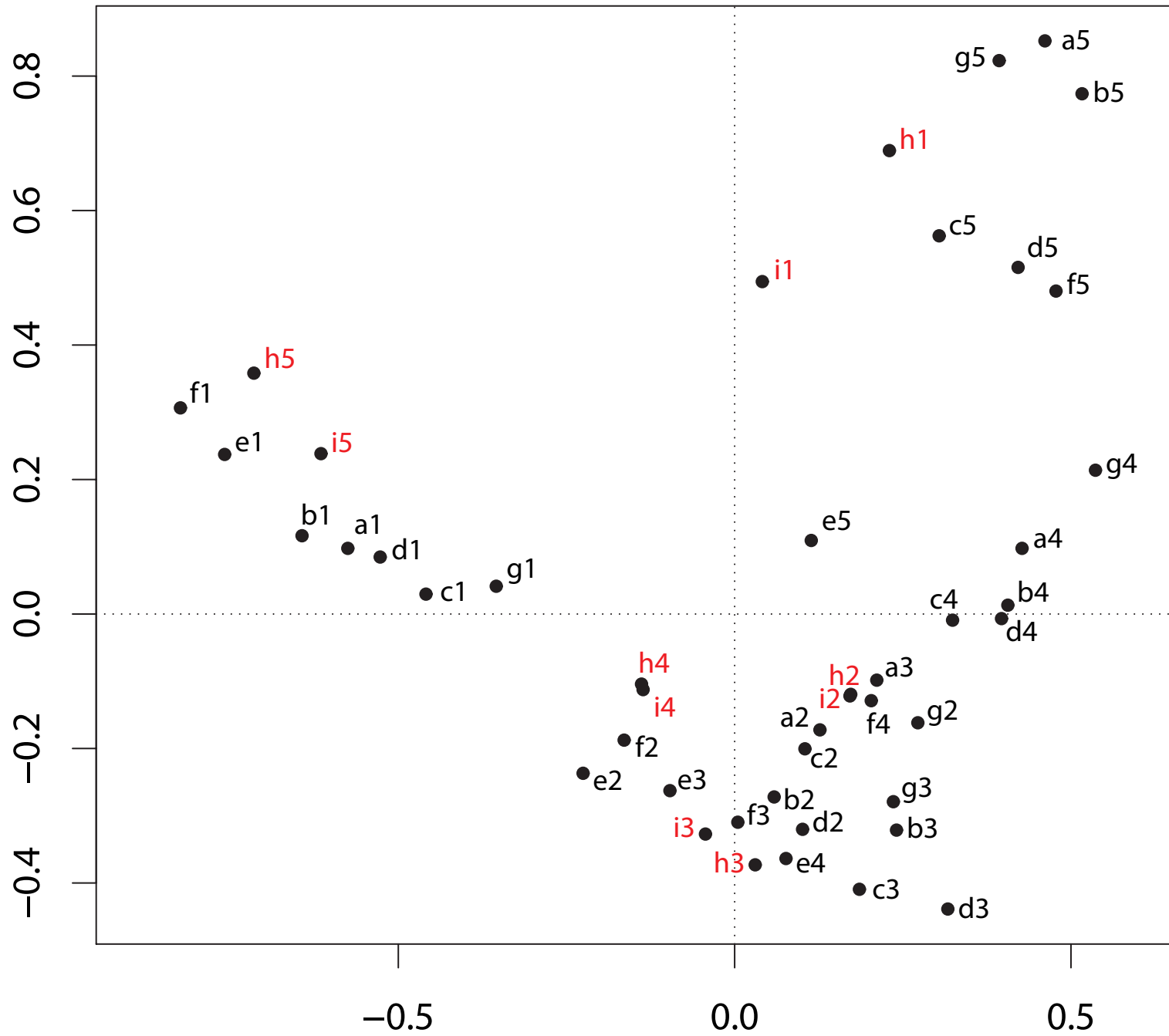
- Best method to see different kinds of methodologically-induced variation, for example, response sets; as well as to distinguish between methodologically-induced and substantive variation
- In MCA and SMCA, similarities between variable categories (or between respondents) are reflected by short (Euclidian) distances, dissimilarities by large distances
- If the quality of data is high, in MCA/SMCA the first dimension should capture mainly substantive variation due to political efficacy and trust, with the second dimension reflecting the horseshoe.
- The items associated with the first dimension should retain their ordinality in this dimension.

- If people did not pay attention to the direction of the questions, the responses to the negatively-formulated items will not conform to an ordinal scale.
- The horseshoe might also appear on the first dimension (large amount of non-substantive variation) or between dimensions 1 and 2 (two-dimensional solution, data might be on high quality).
- If there is a high intercorrelation within the non-substantive responses, in MCA, the first or second dimension will just reflect the difference between substantive and non-substantive responses, in SMCA the non-substantive responses can be excluded without missing any information as it is true in the case in listwise deletion.

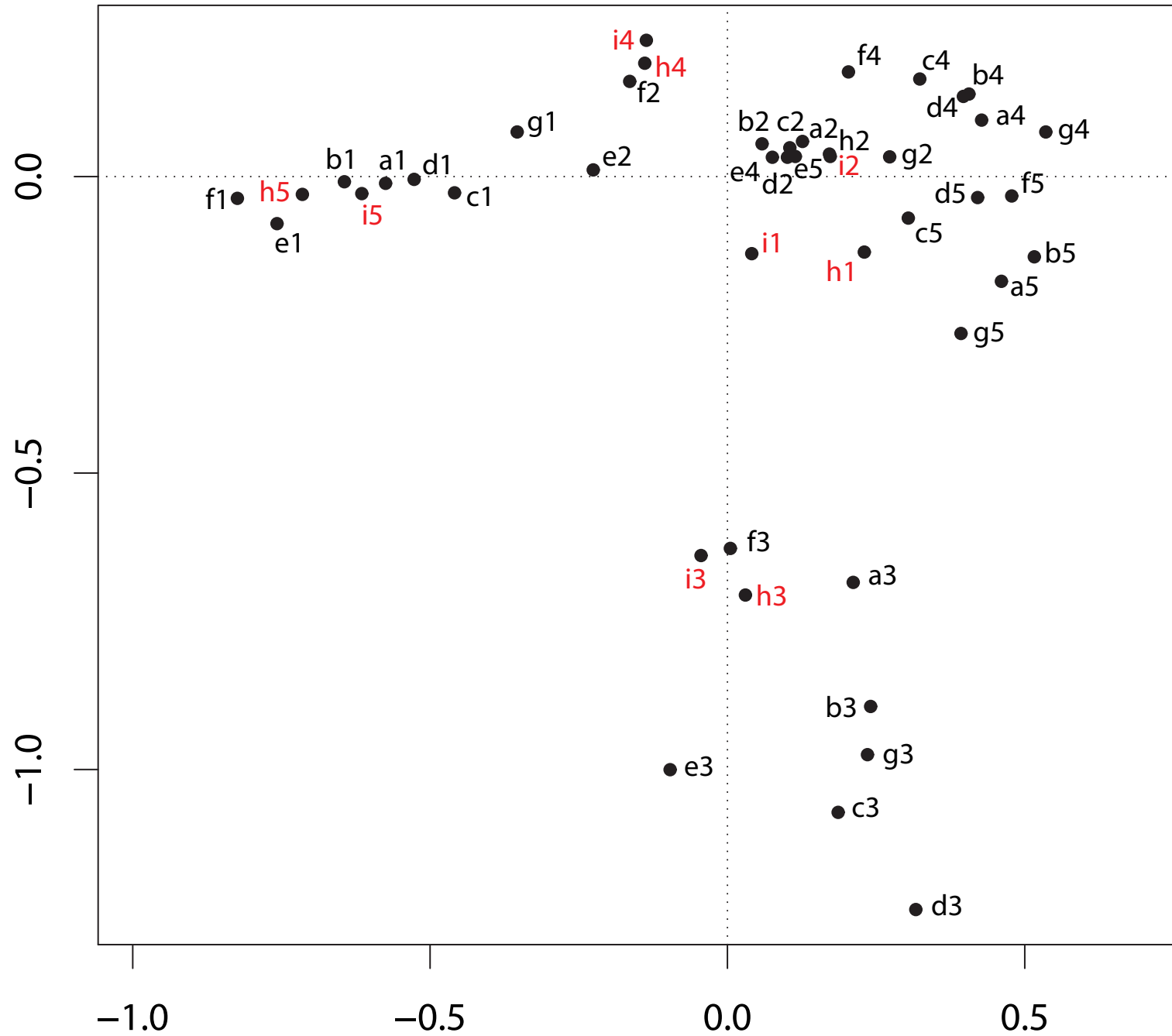
Fedgov, N = 3,377; all respondents



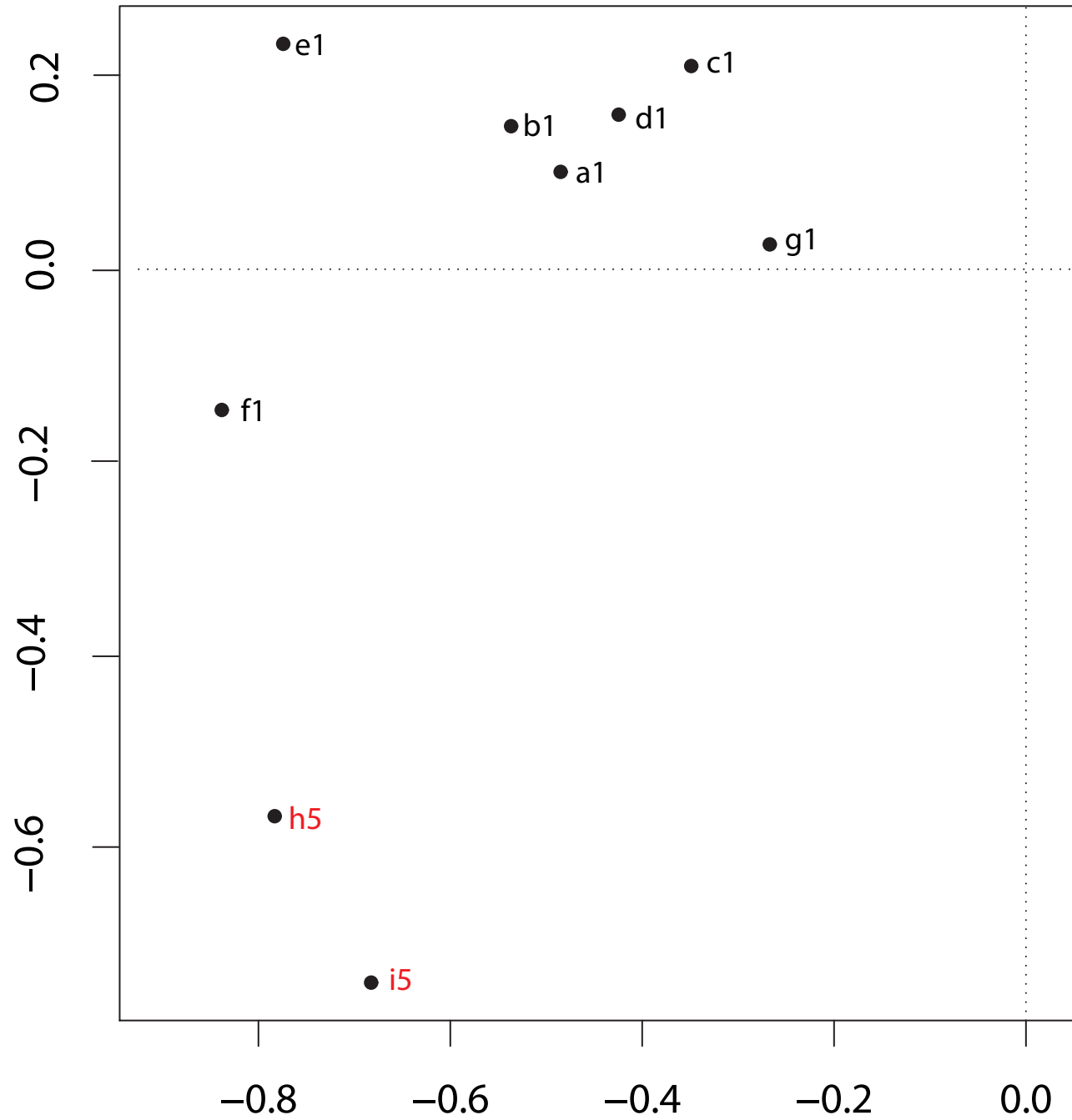
SMCA (1,2,3,4,5), Fedgov, N=3,377



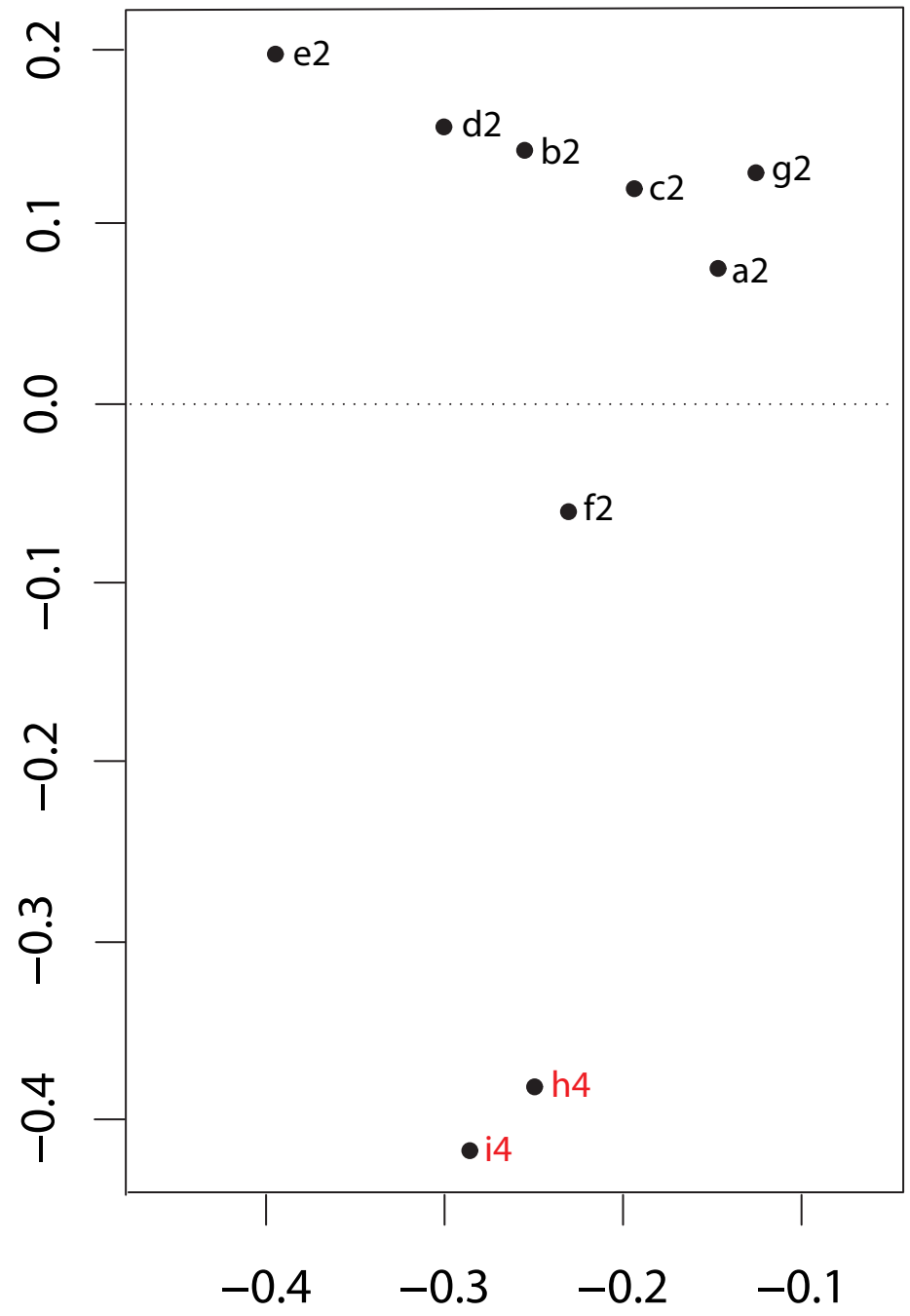
SMCA, D1 vs D3, N = 3,377



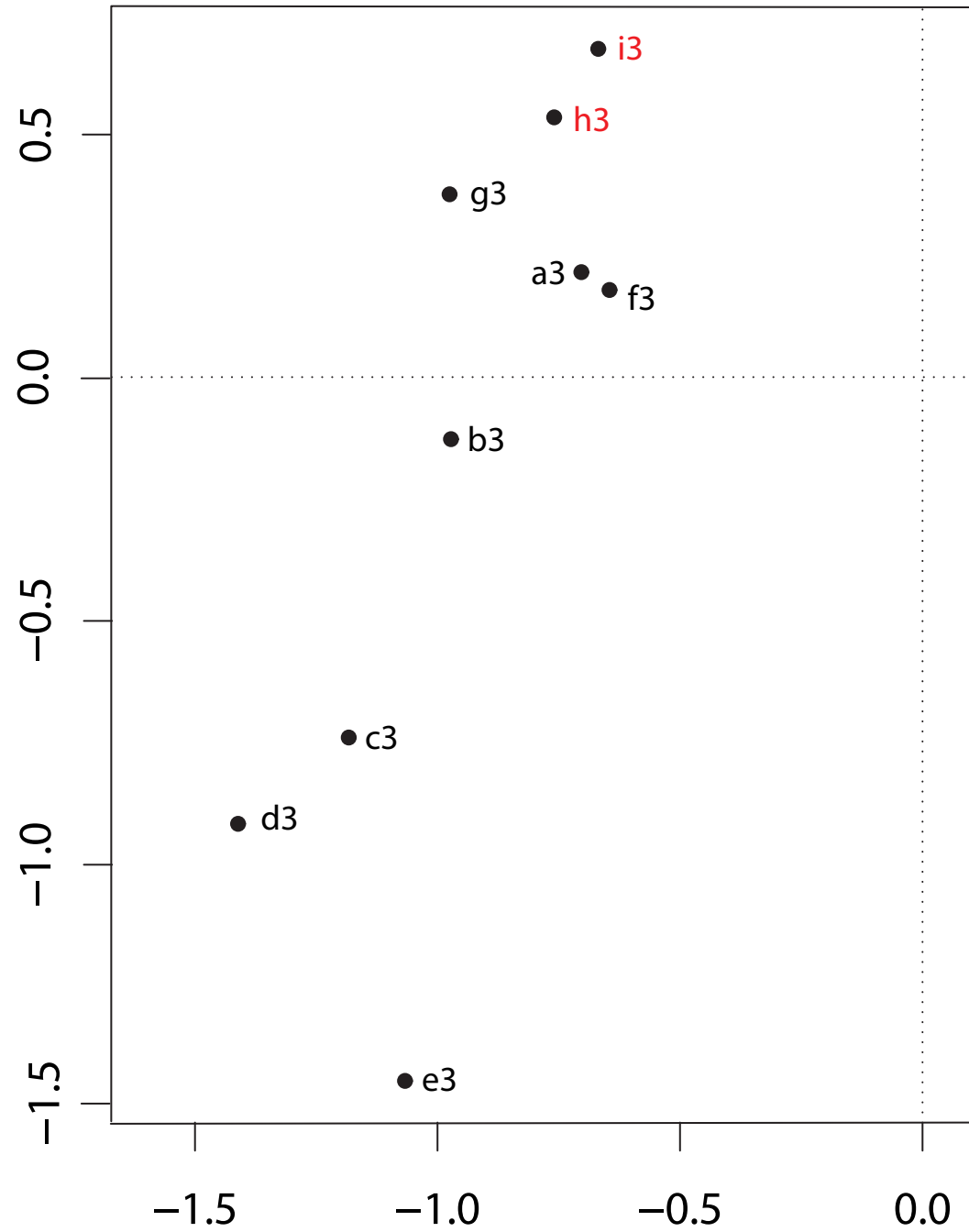
Fedgov, SMCA (1), N=3,377



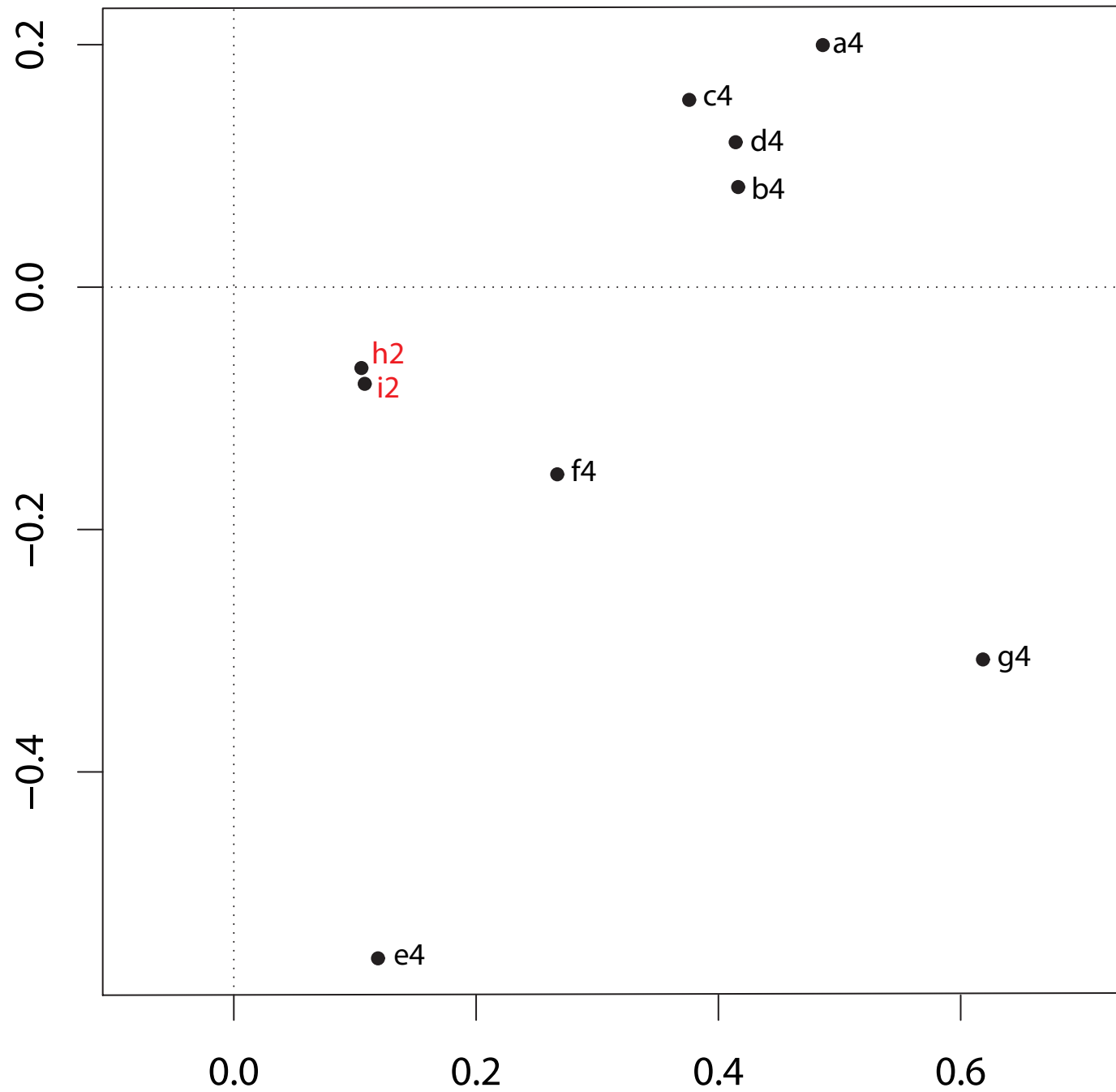
Fedgov, SMCA (2), N=3,377



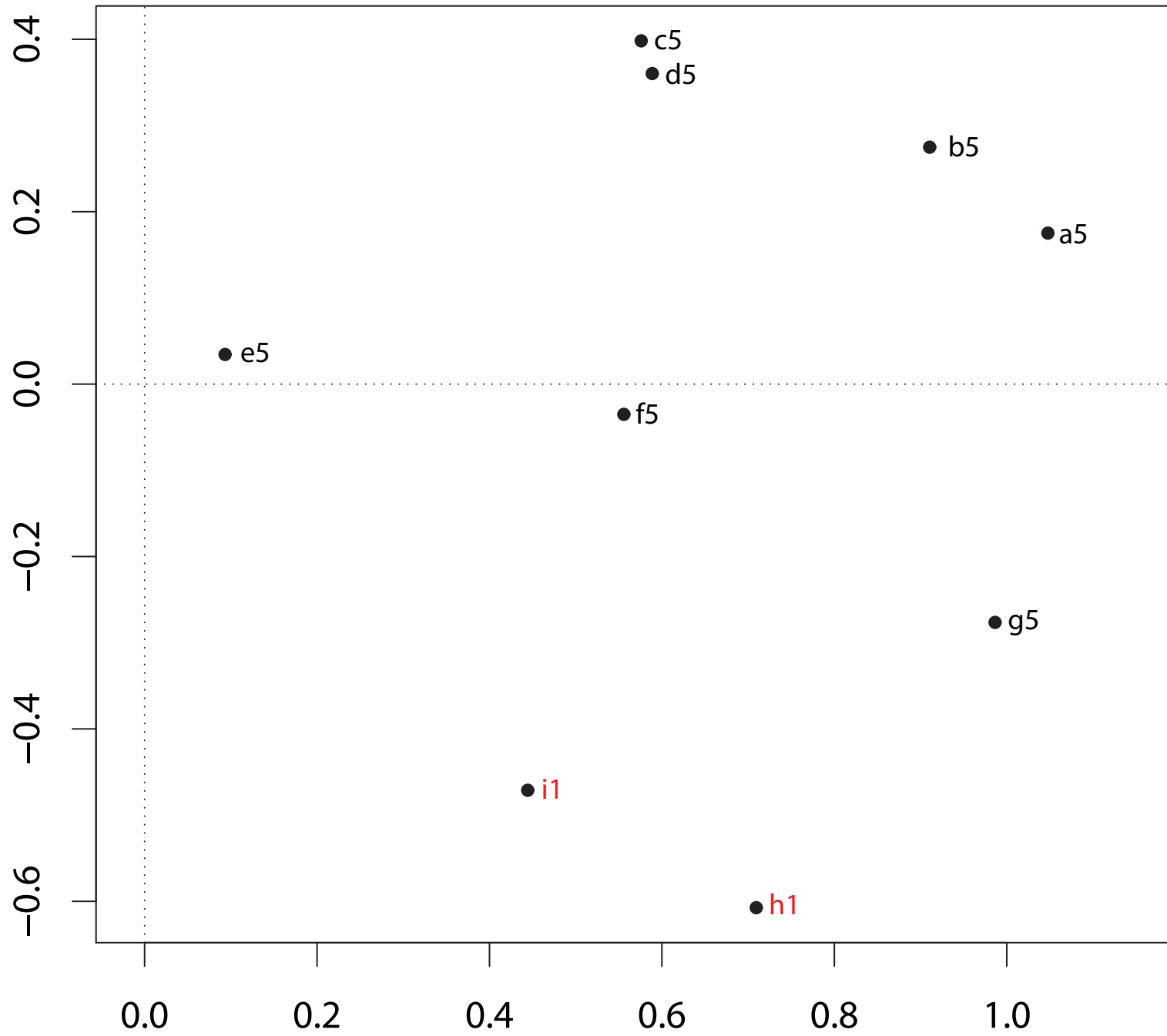
Fedgov, SMCA (3), N=3,377



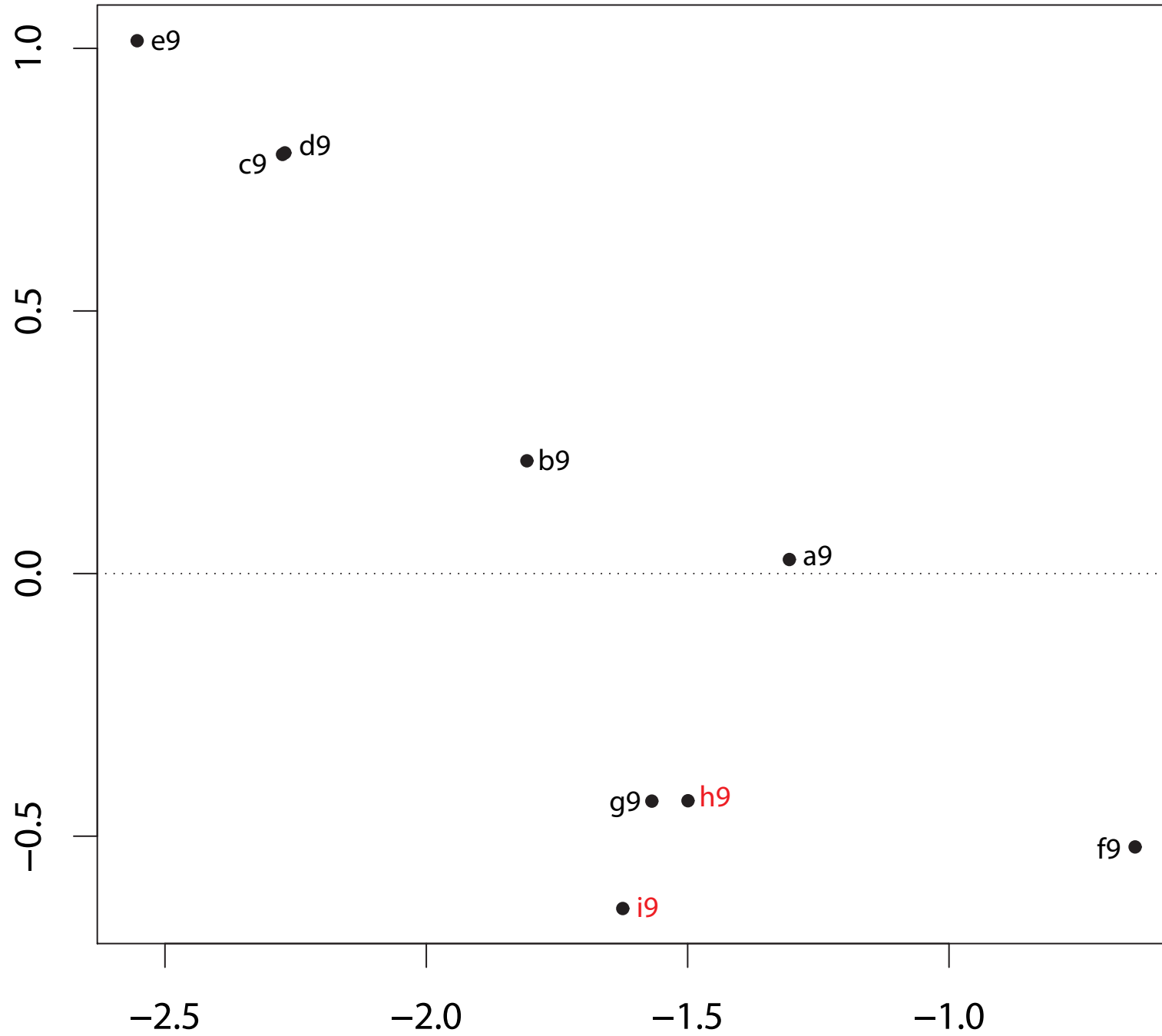
Fedgov, SMCA (4), N=3,377



Fedgov, SMCA (5), N=3,377



Fedgov, SMCA (9), N=3,377



Decomposition of inertia, SMCA, Federal Government

Model	K	Dimension 1		Dimension 2		Total	
		Abs.	In %	Abs.	In %	Abs.	In %
All categories	45	0.1118	15.8	0.1036	14.6	0.7083	100.0
Subset(1,2,3,4,5)	45	0.1107	20.3	0.0625	11.5	0.5441	76.8
Subset(9)	9	0.0929	62.7	0.0117	7.9	0.1481	20.9
Interaction	9	0.0046	57.8	0.0010	12.7	0.0080	1.1
Subset(1,2,4,5)	36	0.1095	25.9	0.0599	14.2	0.4225	59.6
Subset(3,9)	18	0.0934	36.2	0.0327	12.6	0.2583	36.5
Interaction	18	0.0066	47.9	0.0017	12.2	0.0137	1.9
Subset(1)	9	0.0543	56.7	0.0128	13.3	0.0959	13.5
Subset(2)	9	0.0150	24.1	0.0114	18.3	0.0622	8.8
Subset(3)	9	0.0326	30.0	0.0140	12.9	0.1086	15.3
Subset(4)	9	0.0229	32.4	0.0093	13.2	0.0705	10.0
Subset(5)	9	0.0442	45.5	0.0138	14.2	0.0972	13.7
Subset(9)	9	0.0929	62.7	0.0117	7.9	0.1481	20.9

Subset(1): First category items “a” to “g”, last category items “h” and “l”, and so on.

Example: $76.8 + 20.9 + 2 \times 1.1 = 100.0$

Marginals and CatPCA factor loadings on political interest items (N = 3,362 to 3,375)

How often do you ...	Often	Some- times	Seldom	Never	D 1
Read about politics in the newspapers and magazines?	42.1	30.9	17.4	9.6	.623
Watch programs about politics on TV?	32.6	38.5	19.4	9.5	.608
Discuss politics with other people?	23.5	38.7	22.8	15.0	.681
Try to convince friends to vote the same as you?	6.1	12.0	16.9	65.0	.622
Attend a political meeting or rally?	3.8	13.9	19.3	63.0	.780
Contact public officials or politicians?	3.8	16.5	20.6	59.2	.672
Spend time working for a political party?	3.4	9.3	9.7	77.6	.706
Contribute money to a political party or candidate	3.6	10.1	9.6	76.7	.652

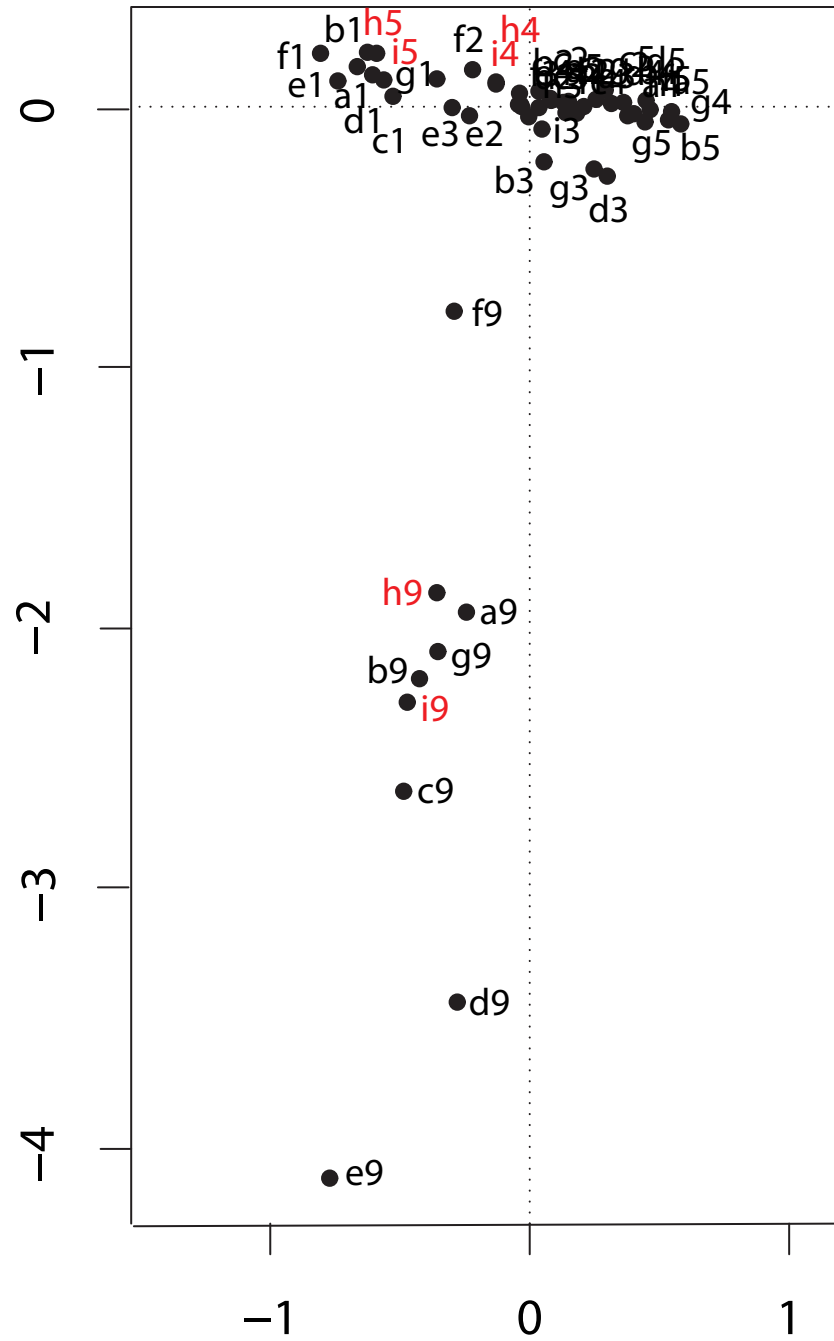
Understanding of questions, subdivision by political interest: First row, low PI, N = 1,935; second row: High PI, N = 1,441

Item	SA	AS	NN	DS	SD	NO	χ^2
a) Generally, those elected to Parliament soon lose touch with the people.	28.1 24.6	44.4 44.8	3.9 3.0	13.5 19.5	3.5 6.5	6.6 1.7	85.2
b) I don't think the (Federal) Government cares much about what people like me think.	30.0 22.7	33.0 32.8	4.1 3.4	22.2 26.9	6.5 12.3	4.2 1.9	69.5
c) Sometimes, (Federal) Politics and Government seem so complicated that a person like me can't really understand what's going on.	38.4 20.7	34.5 31.4	2.8 1.9	15.1 24.4	6.8 20.3	2.3 1.3	249.7
d) People like me don't have any say about what the Government in (Ottawa) does.	38.0 27.1	28.8 27.8	2.8 1.5	16.8 24.3	10.7 18.5	2.9 0.9	111.1
e) So many other people vote in (Federal) elections that it does not matter very much whether I vote or not.	10.4 4.3	12.5 6.5	2.2 1.3	19.2 11.8	53.4 75.4	2.4 0.7	179.2
f) Many people in the (Federal) Government are dishonest.	11.4 9.2	26.0 23.9	11.0 9.0	23.3 26.4	13.4 24.7	15.0 6.8	118.2

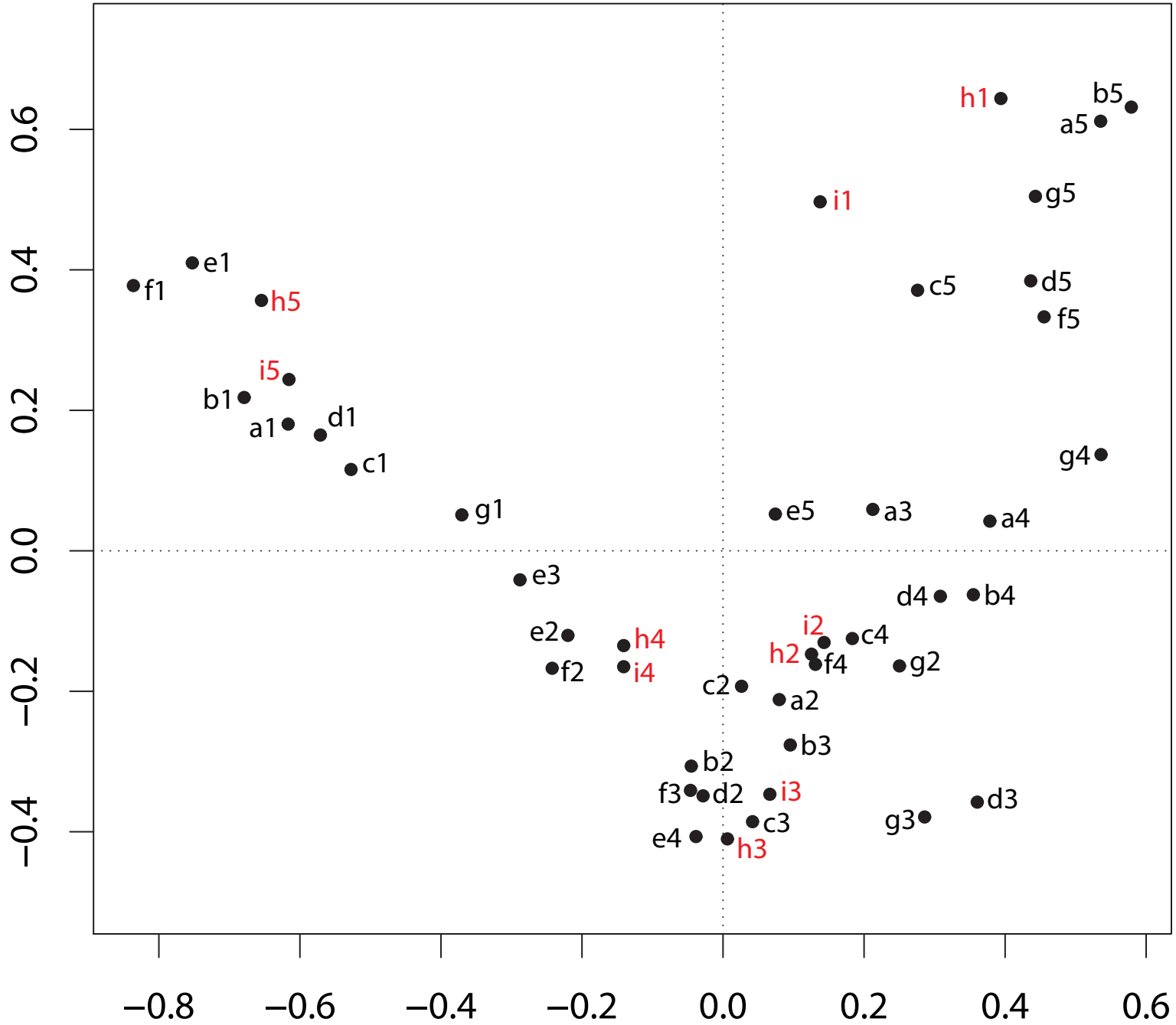
Item	SA	AS	NN	DS	SD	NO	χ^2
g) People in the (Federal) Government waste a lot of the money we pay in taxes.	46.4 46.1	33.3 33.0	4.5 3.0	7.7 10.7	2.5 5.1	5.5 2.0	54.2
h) Most of the time we can trust people in the (Federal) Government to do what is right.	8.8 12.6	47.1 44.6	7.4 4.6	22.4 24.8	8.9 10.8	5.4 2.6	43.9
i) Most of the people running the (Federal) Government are smart people who usually know what they are doing.	14.5 17.8	46.8 43.7	7.1 4.2	19.1 23.5	7.6 9.0	4.9 1.7	51.5

One missing case because one respondents did not answer the political interest items.

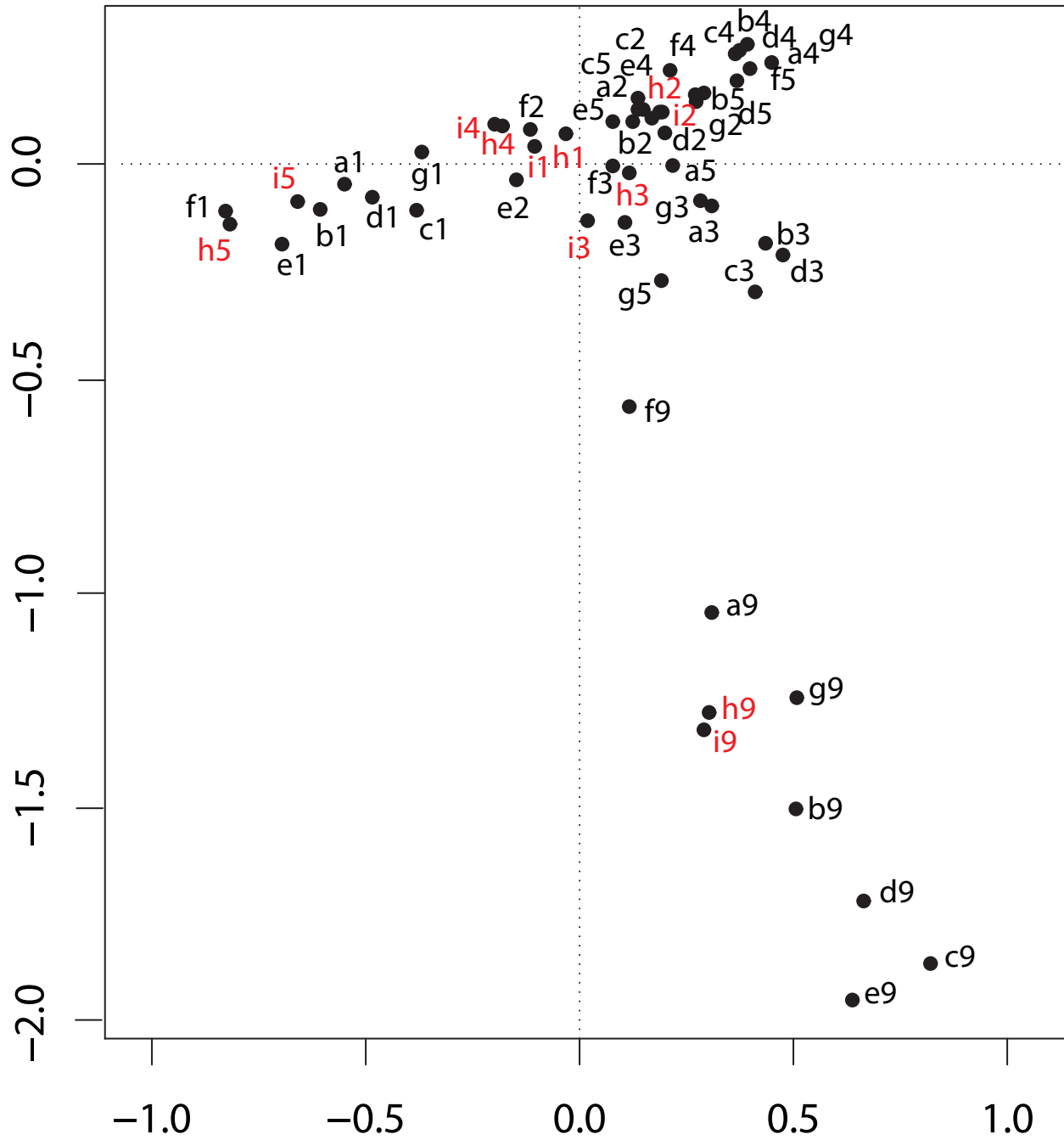
Fedgov, High Political Interest



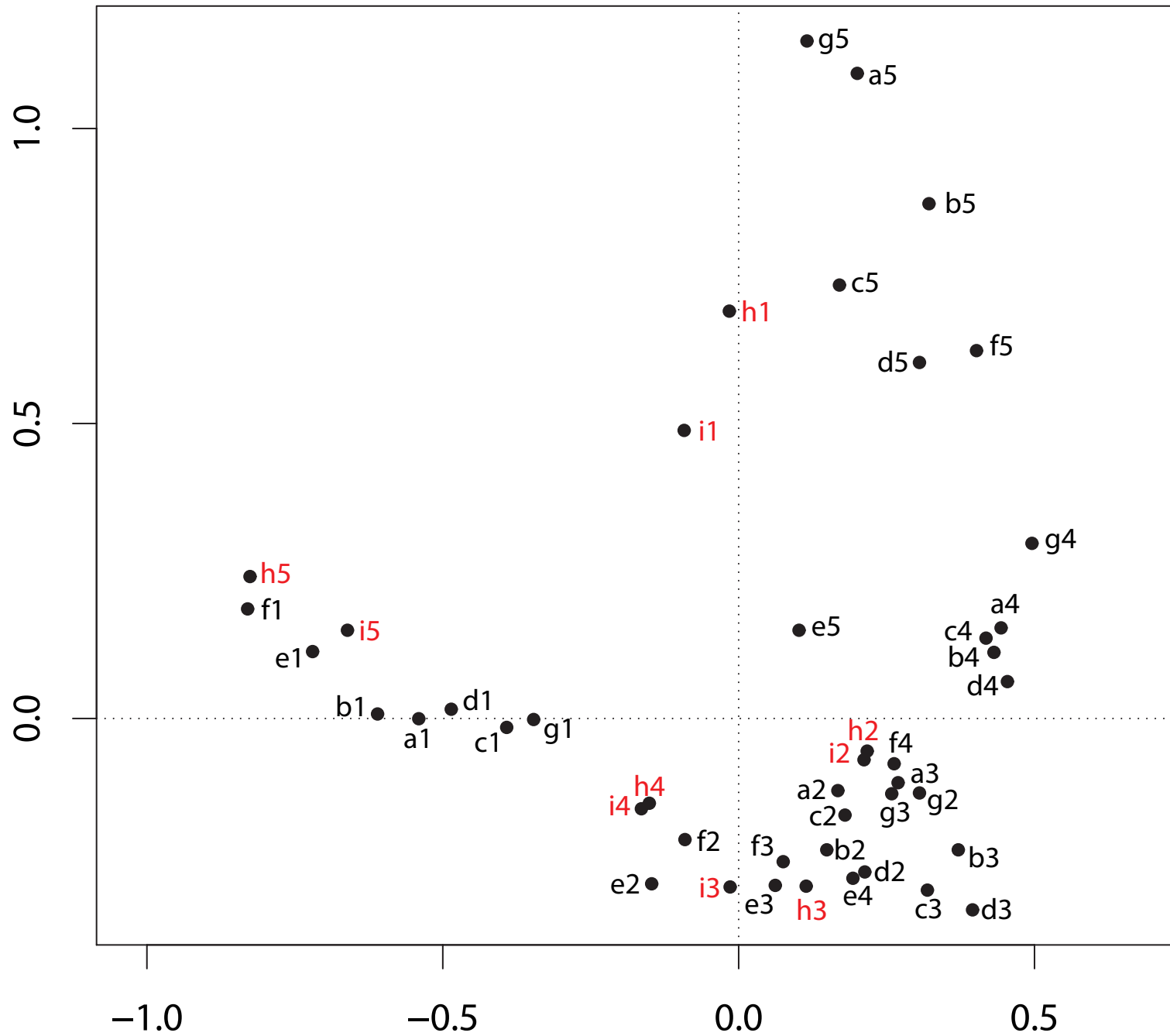
Fedgov, SMCA (1,2,3,4,5), High Political Interest



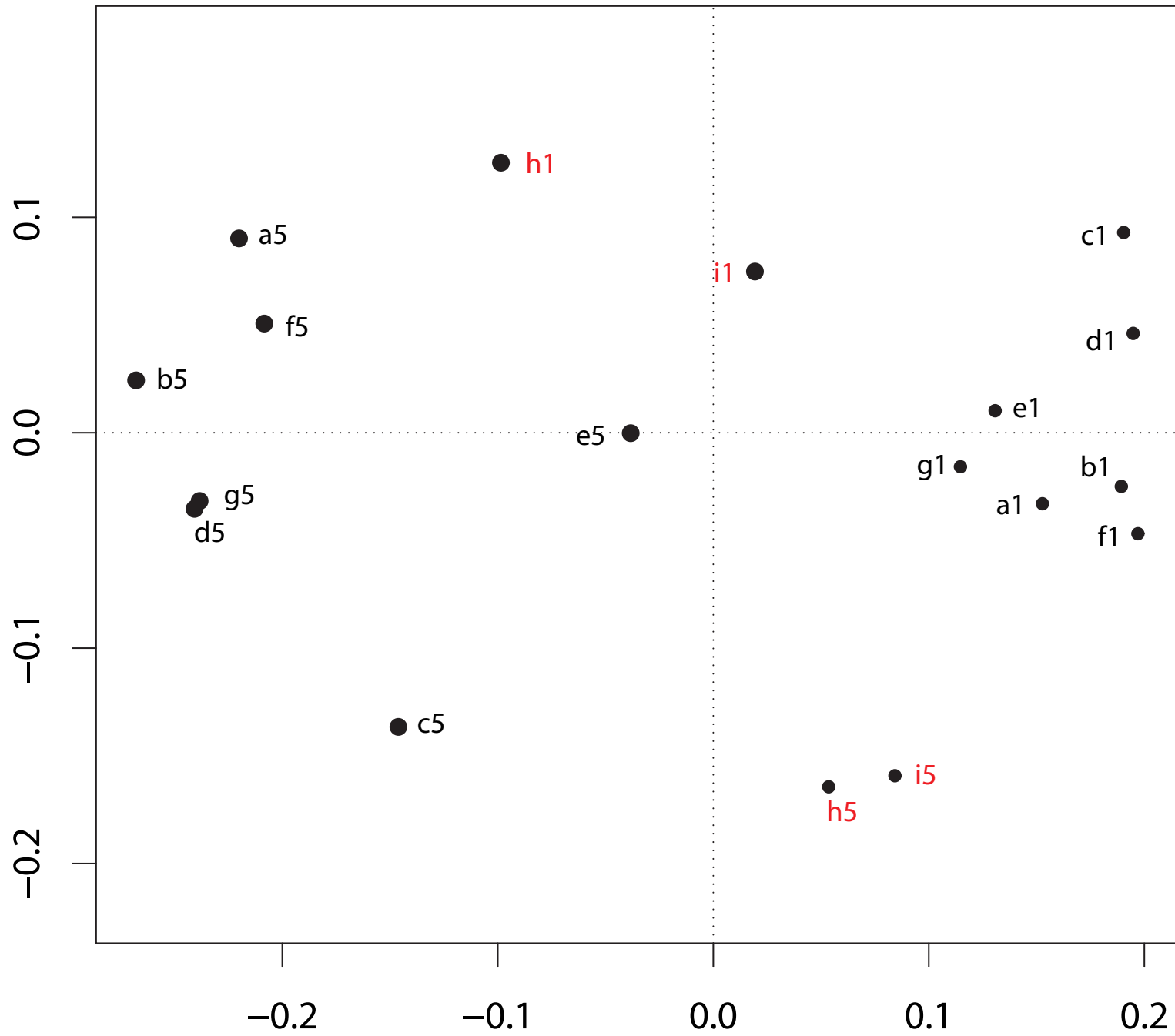
Fedgov, MCA, Low Political Interest



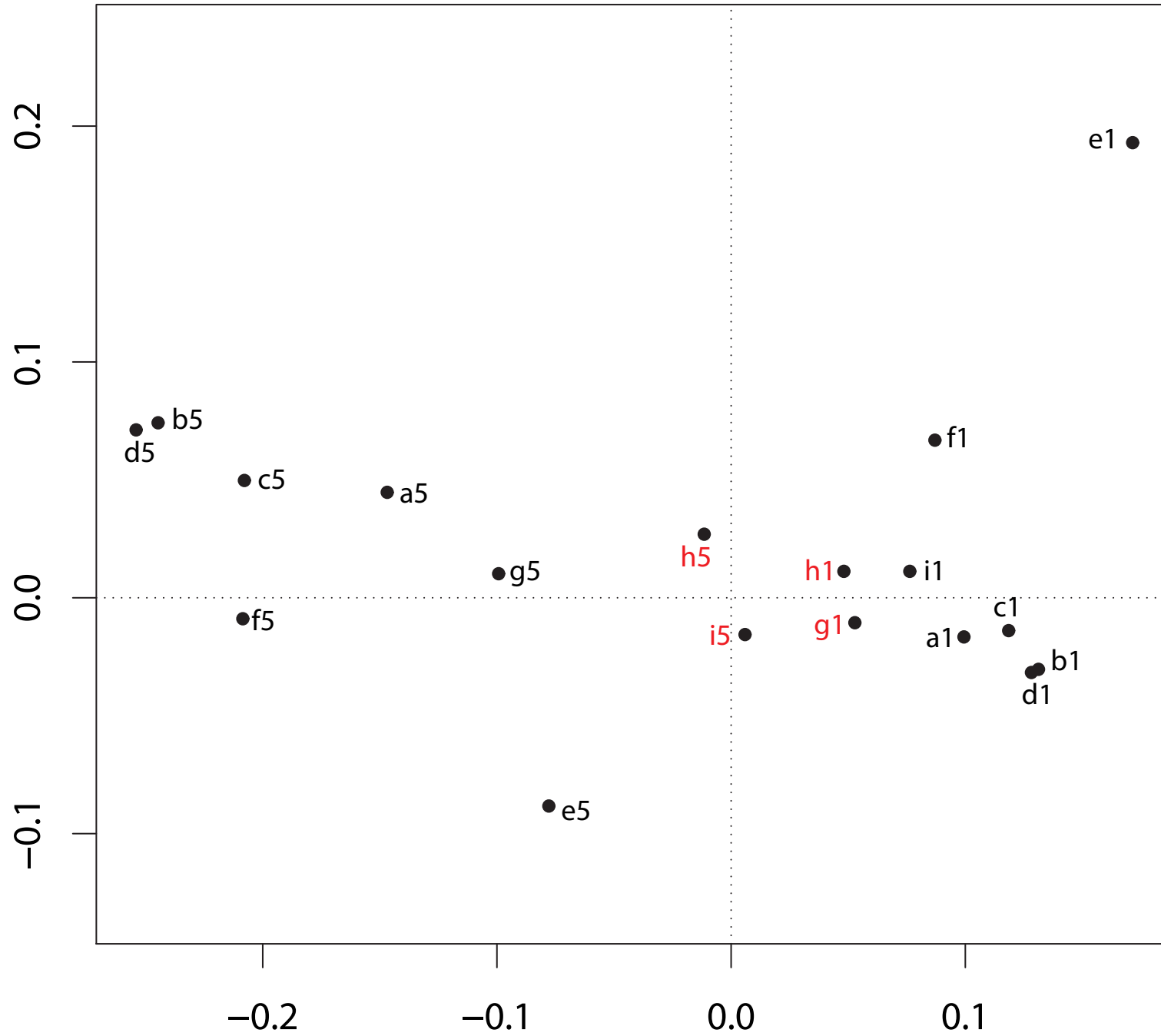
Fedgov, SMCA (1,2,3,4,5), Low Political Interest



Fedgov, high pol. interest, s1 vs s5



Fedgov, low pol. interest, s1 vs s5



Decomposition of inertia, low and high political interest, SMCA of single categories

Model	Low PI, 9 items		High PI, 9 items		Low PI, 7 items		High PI, 7 items	
	Inertia	D1, in %	Inertia	D1, in %	Inertia	D1, in %	Inertia	D1, in %
Subset(1)	0.0915	57.7	0.1010	54.7	0.0629	63.9	0.0733	61.0
Subset(2)	0.0619	24.2	0.0631	24.4	0.0451	31.3	0.0475	29.6
Subset(3)	0.1086	31.6	0.1088	26.2	0.0842	34.1	0.0842	29.0
Subset(4)	0.0741	36.0	0.0664	27.9	0.0655	38.1	0.0576	30.8
Subset(5)	0.0996	40.0	0.0899	46.7	0.0755	44.3	0.0651	49.9
Subset(9)	0.1412	61.0	0.1478	59.4	0.1025	60.6	0.1071	58.7

The Fatigue Effect

The 1984 CNES consists of rather long face-to-face interviews that lasted just over one-and-a-half hours on average. In the first section of the questionnaire, respondents were asked about their attitudes towards the federal government; in the last section, just before the demographic items, the identical questions were asked about the provincial government. This questionnaire architecture is a felicitous feature for assessing what is known as the respondent fatigue effect, which postulates that response quality is a function of the length of the interview, with the response quality decreasing over time, especially near the end of long questionnaires.

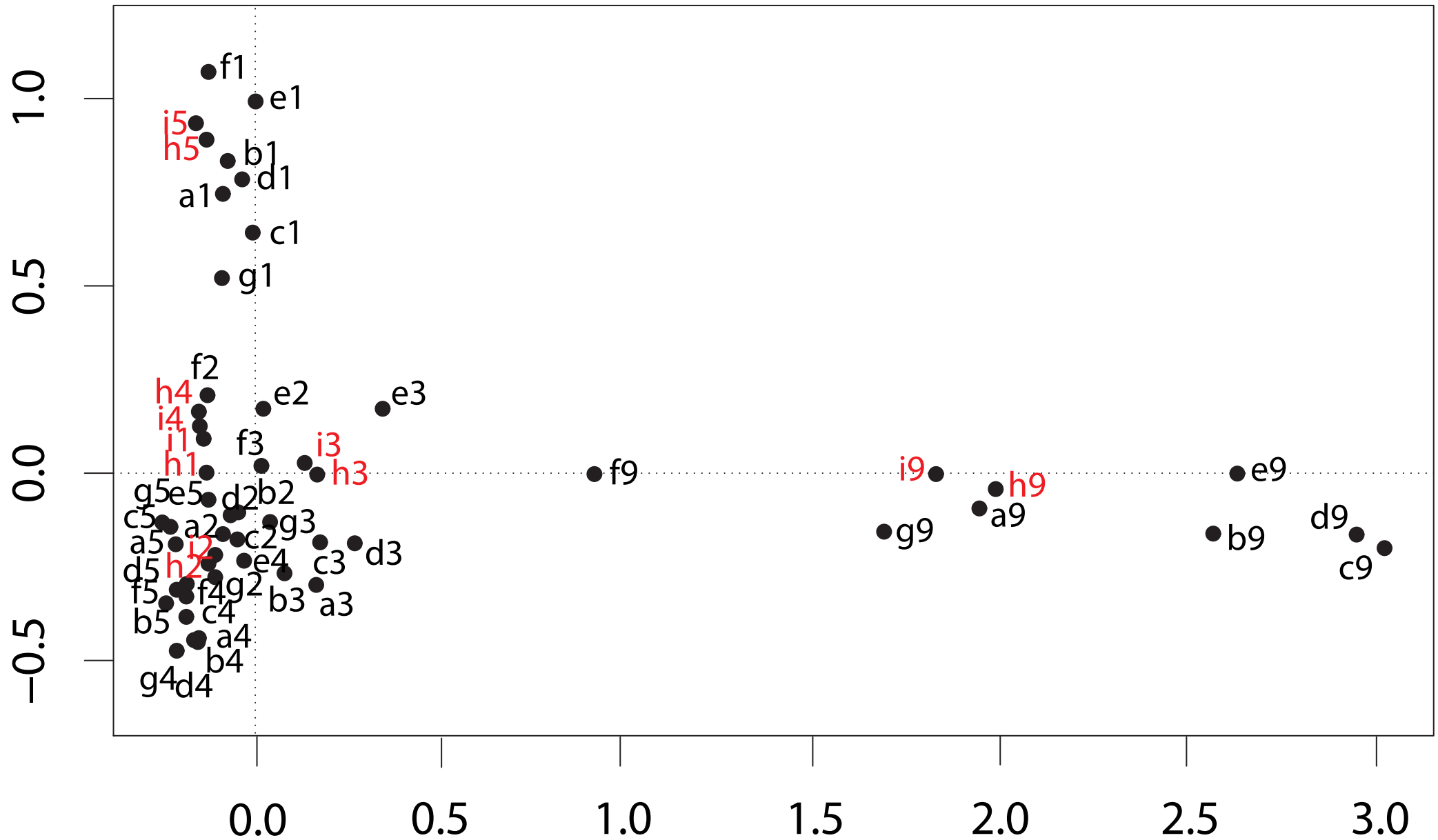
Most respondents expressed equal interest in federal and provincial politics, although approximately 10 percent of the respondents stated that they were more interested in federal than in provincial politics.

Fatigue Effect: First row, Federal Government, N = 3,377; second row:
Provincial Government, N=3,346

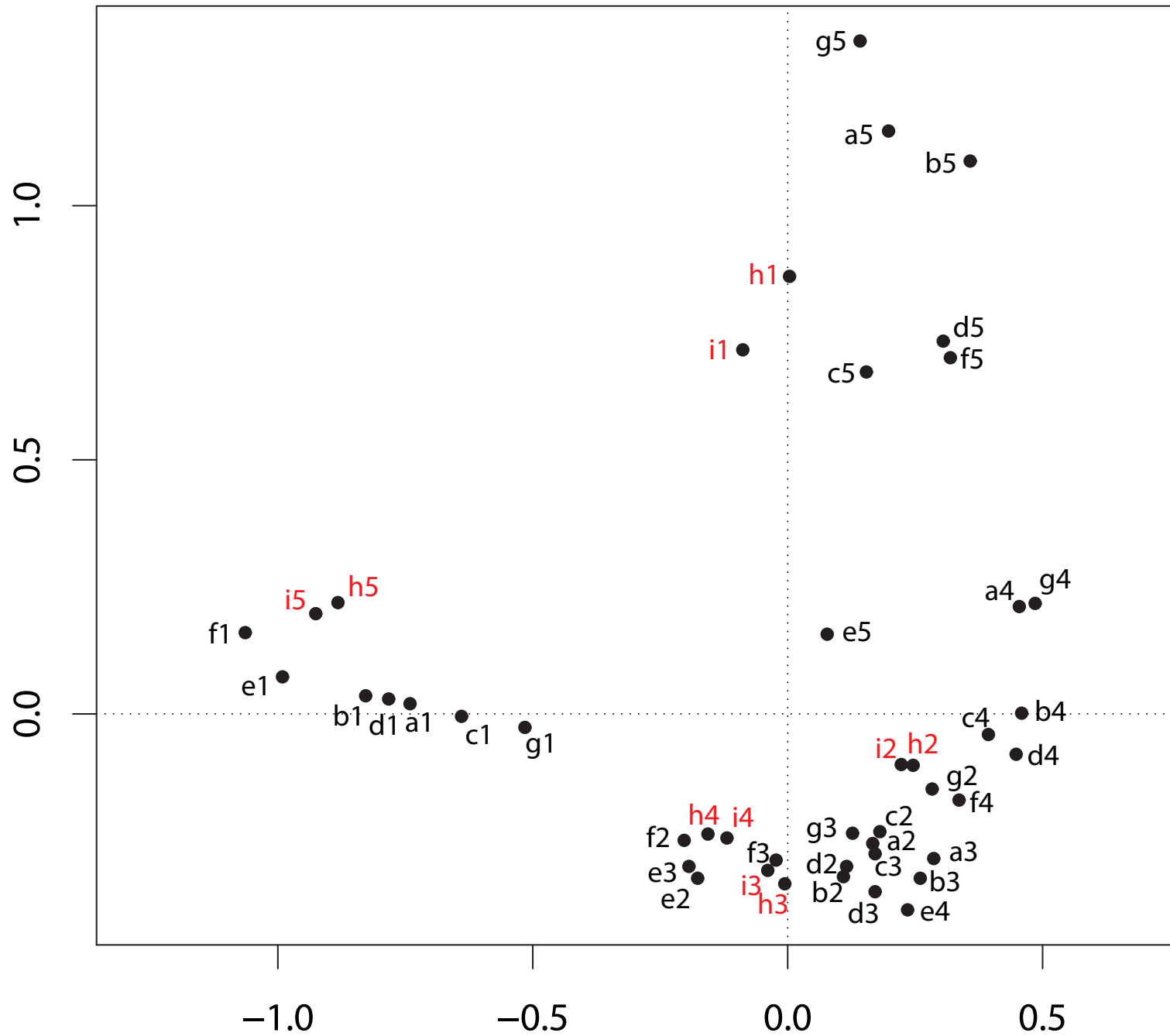
Item	SA	AS	NN	DS	SD	NO
a) Generally, those elected to Parliament soon lose touch with the people.	26.6 24.2	44.5 42.0	3.5 2.7	16.1 19.8	4.8 6.2	4.5 5.1
b) I don't think the (Federal) Government cares much about what people like me think.	26.9 24.7	32.9 31.7	3.8 2.2	24.2 28.4	9.0 9.3	3.2 3.8
c) Sometimes, (Federal) Politics and Government seem so complicated that a person like me can't really understand what's going on.	30.8 25.9	33.1 36.6	2.5 1.6	19.1 18.5	12.6 14.6	1.9 2.8
d) People like me don't have any say about what the Government in (Ottawa) does.	33.4 25.1	28.3 29.7	2.2 2.0	20.0 25.1	14.0 15.1	2.1 3.0
e) So many other people vote in (Federal) elections that it does not matter very much whether I vote or not.	7.8 6.7	9.9 9.0	1.8 1.6	16.0 17.2	62.8 62.5	1.7 2.9
f) Many people in the (Federal) Government are dishonest.	10.5 8.9	25.1 22.9	10.1 9.3	24.6 26.7	18.2 18.3	11.5 14.0
g) People in the (Federal) Government waste a lot of the money we pay in taxes.	46.3 35.2	33.2 39.3	3.9 3.7	9.0 11.7	3.6 3.8	4.1 6.3

Item	SA	AS	NN	DS	SD	NO
h) Most of the time we can trust people in the (Federal) Government to do what is right.	10.4	46.0	6.2	23.5	9.7	4.2
	11.0	49.2	5.7	18.4	10.2	5.4
i) Most of the people running the (Federal) Government are smart people who usually know what they are doing.	15.9	45.5	5.9	21.0	8.2	3.6
	14.1	49.3	5.1	18.2	7.5	5.7

Provgov, MCA, all respondents



Provgov, SMCA (1,2,3,4,5), all respondents



Decomposition of inertia, SMCA, Provincial and Federal Government

Model	K	Provincial Government				Federal Government			
		D1	Total	D1	Total	D1	Total	D1	Total
		Abs.	In %	Abs.	In %	Abs.	In %	Abs.	In %
All categories	45	0.2222	24.9	0.8914	100.0	0.1118	15.8	0.7083	100.0
Subset(1-5)	45	0.1568	24.9	0.6309	70.8	0.1107	20.3	0.5441	76.8
Subset(9)	9	0.1972	85.0	0.2320	26.0	0.0929	62.7	0.1481	20.9
Interaction	9	0.0113	79.3	0.0143	1.6	0.0046	57.8	0.0080	1.1
Subset(1,2,4,5)	36	0.1559	31.2	0.4989	56.0	0.1095	25.9	0.4225	59.6
Subset(3,9)	18	0.1976	56.0	0.3531	39.6	0.0934	36.2	0.2583	36.5
Interaction	18	0.0137	69.6	0.0197	2.2	0.0066	47.9	0.0137	1.9
Subset(1)	9	0.0865	68.2	0.1268	14.2	0.0543	56.7	0.0959	13.5
Subset(2)	9	0.0239	34.7	0.0689	7.7	0.0150	24.1	0.0622	8.8
Subset(3)	9	0.0487	40.7	0.1197	13.4	0.0326	30.0	0.1086	15.3
Subset(4)	9	0.0340	46.1	0.0737	8.3	0.0229	32.4	0.0705	10.0
Subset(5)	9	0.0700	60.9	0.1150	12.9	0.0442	45.5	0.0972	13.7
Subset(9)	9	0.1972	85.0	0.2320	26.0	0.0929	62.7	0.1481	20.9

Decomposition of inertia, SMCA, Provincial and Federal Government

Model	K	Provincial Government				Federal Government			
		D1		Total		D1		Total	
		Abs.	In %	Abs.	In %	Abs.	In %	Abs.	In %
All categories	45	0.2222	24.9	0.8914	100.0	0.1118	15.8	0.7083	100.0
Subset(1-5)	45	0.1568	24.9	0.6309	70.8	0.1107	20.3	0.5441	76.8
Subset(9)	9	0.1972	85.0	0.2320	26.0	0.0929	62.7	0.1481	20.9
Interaction	9	0.0113	79.3	0.0143	1.6	0.0046	57.8	0.0080	1.1
Subset(1,2,4,5)	36	0.1559	31.2	0.4989	56.0	0.1095	25.9	0.4225	59.6
Subset(3,9)	18	0.1976	56.0	0.3531	39.6	0.0934	36.2	0.2583	36.5
Interaction	18	0.0137	69.6	0.0197	2.2	0.0066	47.9	0.0137	1.9
Subset(1)	9	0.0865	68.2	0.1268	14.2	0.0543	56.7	0.0959	13.5
Subset(2)	9	0.0239	34.7	0.0689	7.7	0.0150	24.1	0.0622	8.8
Subset(3)	9	0.0487	40.7	0.1197	13.4	0.0326	30.0	0.1086	15.3
Subset(4)	9	0.0340	46.1	0.0737	8.3	0.0229	32.4	0.0705	10.0
Subset(5)	9	0.0700	60.9	0.1150	12.9	0.0442	45.5	0.0972	13.7
Subset(9)	9	0.1972	85.0	0.2320	26.0	0.0929	62.7	0.1481	20.9

Decomposition of inertia, SMCA, Provincial and Federal Government

Model	K	Provincial Government				Federal Government			
		D1		Total		D1		Total	
		Abs.	In %	Abs.	In %	Abs.	In %	Abs.	In %
All categories	45	0.2222	24.9	0.8914	100.0	0.1118	15.8	0.7083	100.0
Subset(1-5)	45	0.1568	24.9	0.6309	70.8	0.1107	20.3	0.5441	76.8
Subset(9)	9	0.1972	85.0	0.2320	26.0	0.0929	62.7	0.1481	20.9
Interaction	9	0.0113	79.3	0.0143	1.6	0.0046	57.8	0.0080	1.1
Subset(1,2,4,5)	36	0.1559	31.2	0.4989	56.0	0.1095	25.9	0.4225	59.6
Subset(3,9)	18	0.1976	56.0	0.3531	39.6	0.0934	36.2	0.2583	36.5
Interaction	18	0.0137	69.6	0.0197	2.2	0.0066	47.9	0.0137	1.9
Subset(1)	9	0.0865	68.2	0.1268	14.2	0.0543	56.7	0.0959	13.5
Subset(2)	9	0.0239	34.7	0.0689	7.7	0.0150	24.1	0.0622	8.8
Subset(3)	9	0.0487	40.7	0.1197	13.4	0.0326	30.0	0.1086	15.3
Subset(4)	9	0.0340	46.1	0.0737	8.3	0.0229	32.4	0.0705	10.0
Subset(5)	9	0.0700	60.9	0.1150	12.9	0.0442	45.5	0.0972	13.7
Subset(9)	9	0.1972	85.0	0.2320	26.0	0.0929	62.7	0.1481	20.9

Conclusion

- Using scaling methods such as PCA and MCA allows for detecting same responses pattern as well as duplicates since they provide with identical factor scores.
- For detecting duplicates, the factor scores are meaningless; the variables are independent from each other.
- While same response pattern are probably caused by interviews who simplified their task, duplicates are probably caused by institutes – copy and paste is also a way task simplification.
- Both, same response patterns and duplicates are uncorrelated with the number of missing values.
- The percentages of same response pattern and duplicates can be used as a method for assigning the quality of the survey data.

- MCA can be used for screening the data, for example, to see if questions in all countries asked for in the same direction, to estimate the number of dimensions in a set of items, and to check if the structure of data is meaningful.
- In contrast to other methods, SMCA shows the structures of single responses *without* losing any information.
- Applying SMCA, it is also possible to exclude missing data (item non-response) without using “listwise deletion” or some kind of imputation techniques.
- Comparing the inertias and the explained variances from SMCA of dimension 1 indicates differences in the amount of method-induced variation, in the given examples acquiescence and the understanding of questions.