RadoNorm

Managing risks from radon and NORM

On-line, interactive training course
The art of public opinion survey analysis: Surveying the public on Radon & NORM

April 2021

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

RadoNorm
Managing risks from radon and NORM

Day 4: Analysis of survey data: Exploratory Techniques
29 April 2021
https://zoom.us/j/92190920610?pwd=bGNlcmxUcSs3aTBVeFpOT2l4eWFFQT09

Time (CET)	**Activity**	**Lead**
09:30-10:30	Exploratory measurement techniques; reliability	Peter
10:30-10:45	*Break (15 minutes)*	
10:45-12:00	Factor analysis, cluster analysis	Peter
12:00-13:30	*Break (1 hour 30 minutes)*	
13:30-13:35	Instructions for individual and group work	Plenary
13:35-15:45	Group 1: Testing latent constructs of own nomological network (SPSS)	Peter, Melisa
	Group 2: Evaluating national reports	Tanja, Peter
15:45-16:00	Summary/Quiz	

RadoNorm

Managing risks from radon and NORM

Exploratory measurement techniques: Factor analysis, reliability analysis, and cluster analysis

Peter Thijssen

Thursday 29 April 2021

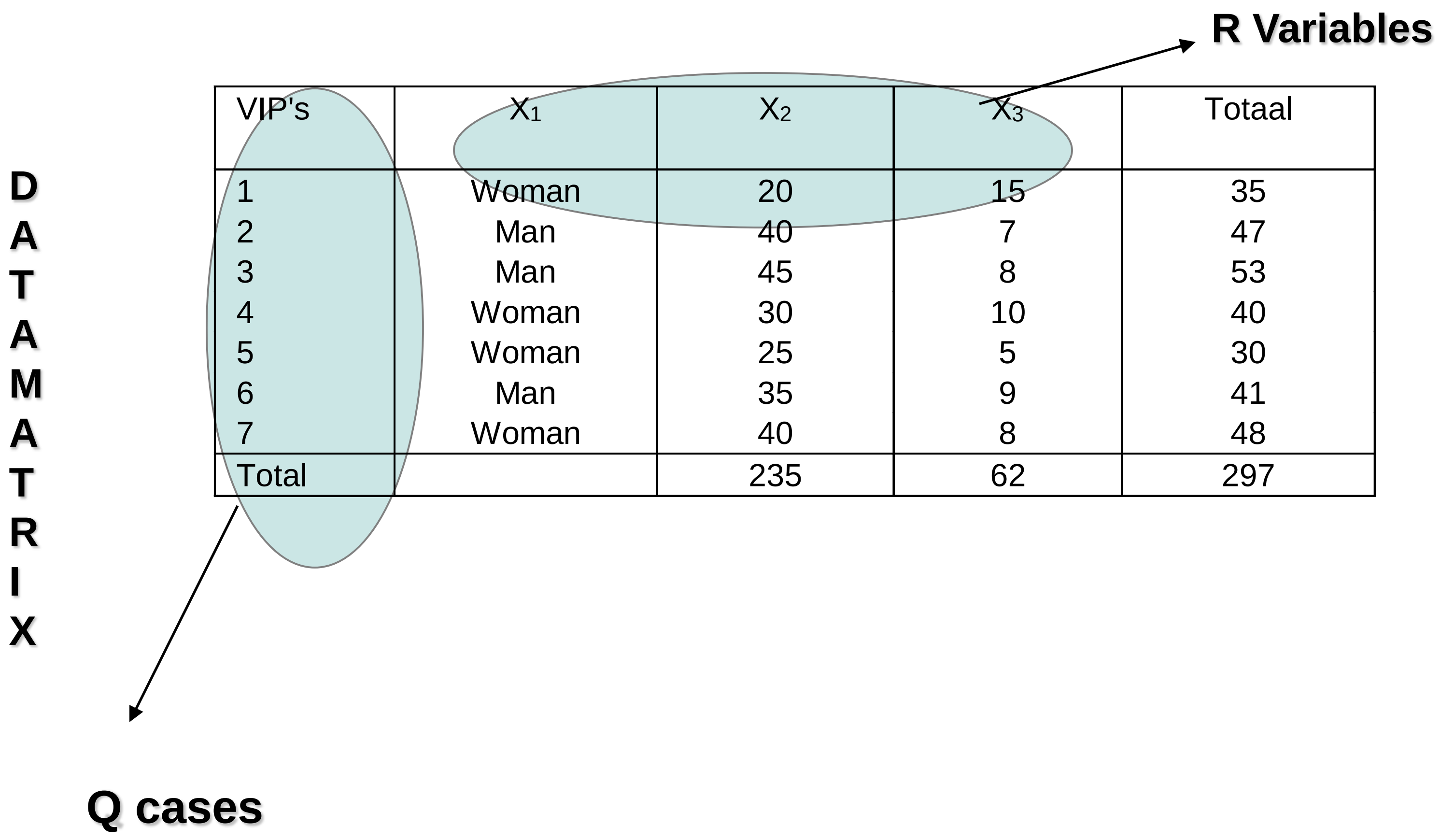
This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Factor analysis

In order to test the validity of indicators as measures of latent constructs

This project has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 900009.

Q- versus R-factor analysis



R Variables

D
A
T
A
M
A
T
R
I
X

VIP's	X_1	X_2	X_3	Totaal
1	Woman	20	15	35
2	Man	40	7	47
3	Man	45	8	53
4	Woman	30	10	40
5	Woman	25	5	30
6	Man	35	9	41
7	Woman	40	8	48
Total		235	62	297

Q cases

Dimensionality of a set of indicators
Factor analysis (FA)

Looking for underlying (latent) common meaning contents that are available in a number of observed variables

For example “efficacy scale” -> to reduce the risk of natural radiation

- Internal locus of control
- External locus of control

2 families of FA

1. Principal components analysis

2. Factor analysis

Principal Component Analysis (PCA)

WHAT?

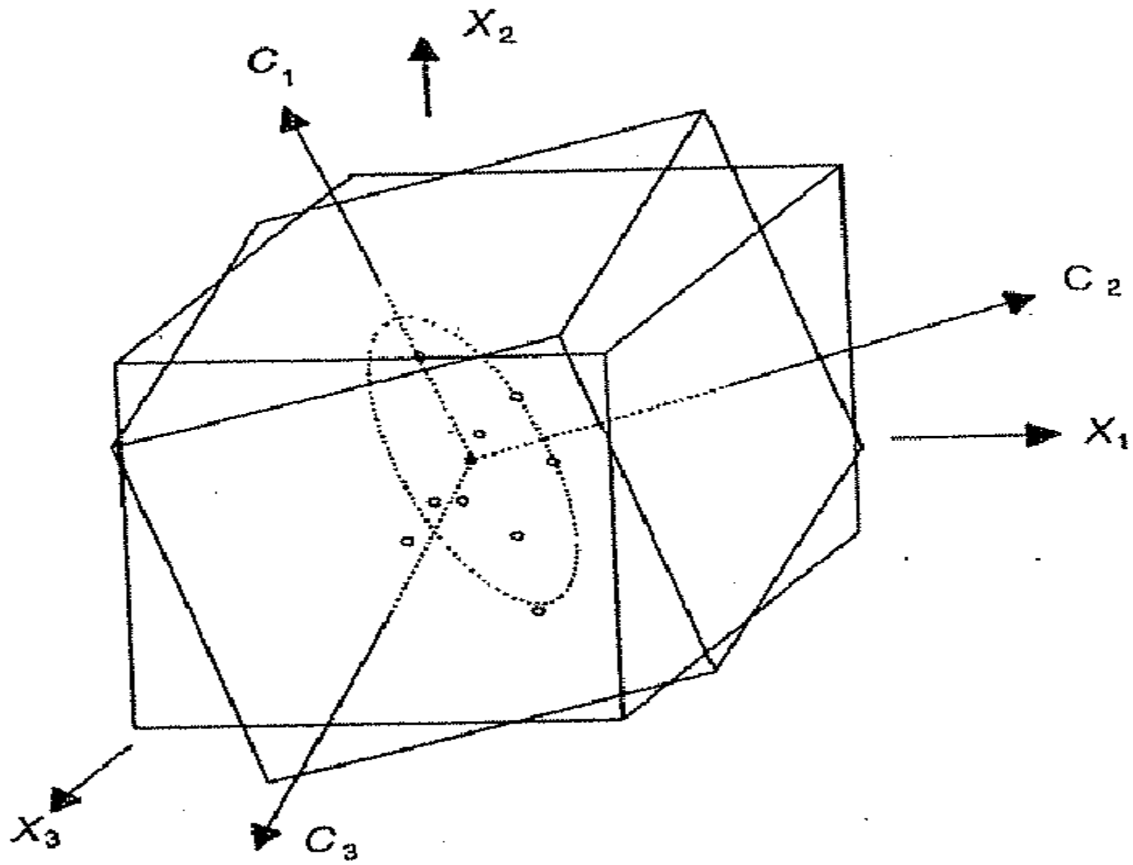
1) Focus on total variance

2) Looking for the same number of components (‘latent constructs’) than there are observed variables, given that:

- Components are *orthogonal* (uncorrelated)
- The components sequentially extract the *maximal amount of variance* from the variables (=principal axis method)

(3 => Selecting the necessary components, in search of a ‘*simple structure*’)
if step 3 is included PCA ~ PFA

Geometric interpretation of PCA



Principal Factor Analysis (PFA)

Crucial elements

1) Partitioning the initial item variance in a common component, specific component and an error component.

2) Looking for a limited number of factors (‘latent constructs’) that explain the *common variance* as good as possible. These factors can be *orthogonal* (uncorrelated) or *oblique* (correlated).

3) => Selecting the necessary factors, in search of a ‘*simple structure*’

FA - Terminology

Factor loading a_{ij} (matrix A: Factor pattern)

$$z_1 = a_{11}f_1 + a_{12}f_2 + a_{13}f_3 (+ e_1)$$

Regression coefficients in a model with a standardized observed variable as dependent and the factors as independents.

Factor (regression)scores u_{ij}

$$f_1 = u_{11}z_1 + u_{12}z_2 + u_{13}z_3$$

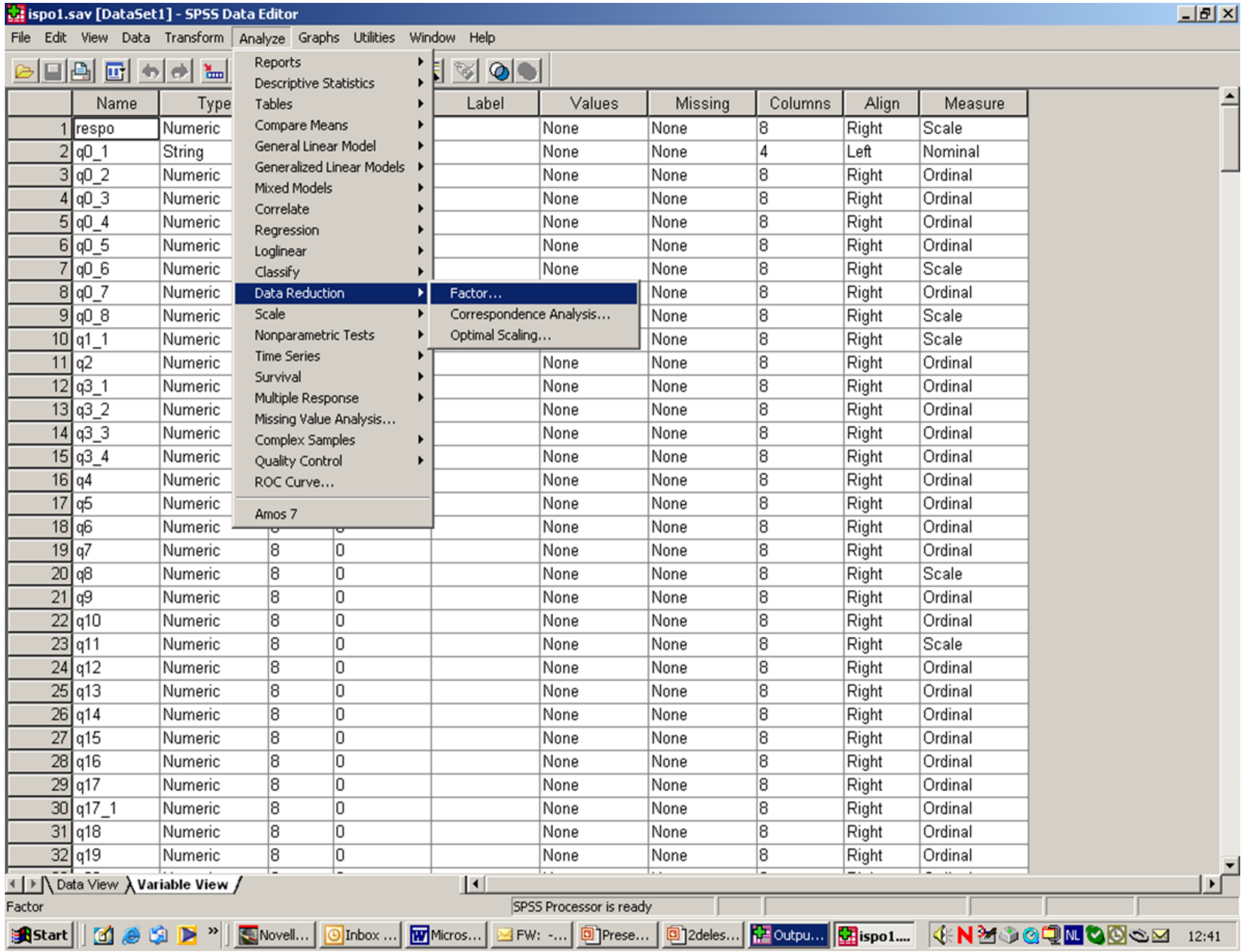
Regression coefficients in a model with a standardized factor as dependent and the standardized observed variables as independents.

Eigenvalue: variance of the projections of each observations on a certain factor; sum of the squared factor loadings

FA – How many factors in the simple structure?

Communality: (common variance) How many percent of the variance of a variable is explained by a (number of) factor(s)

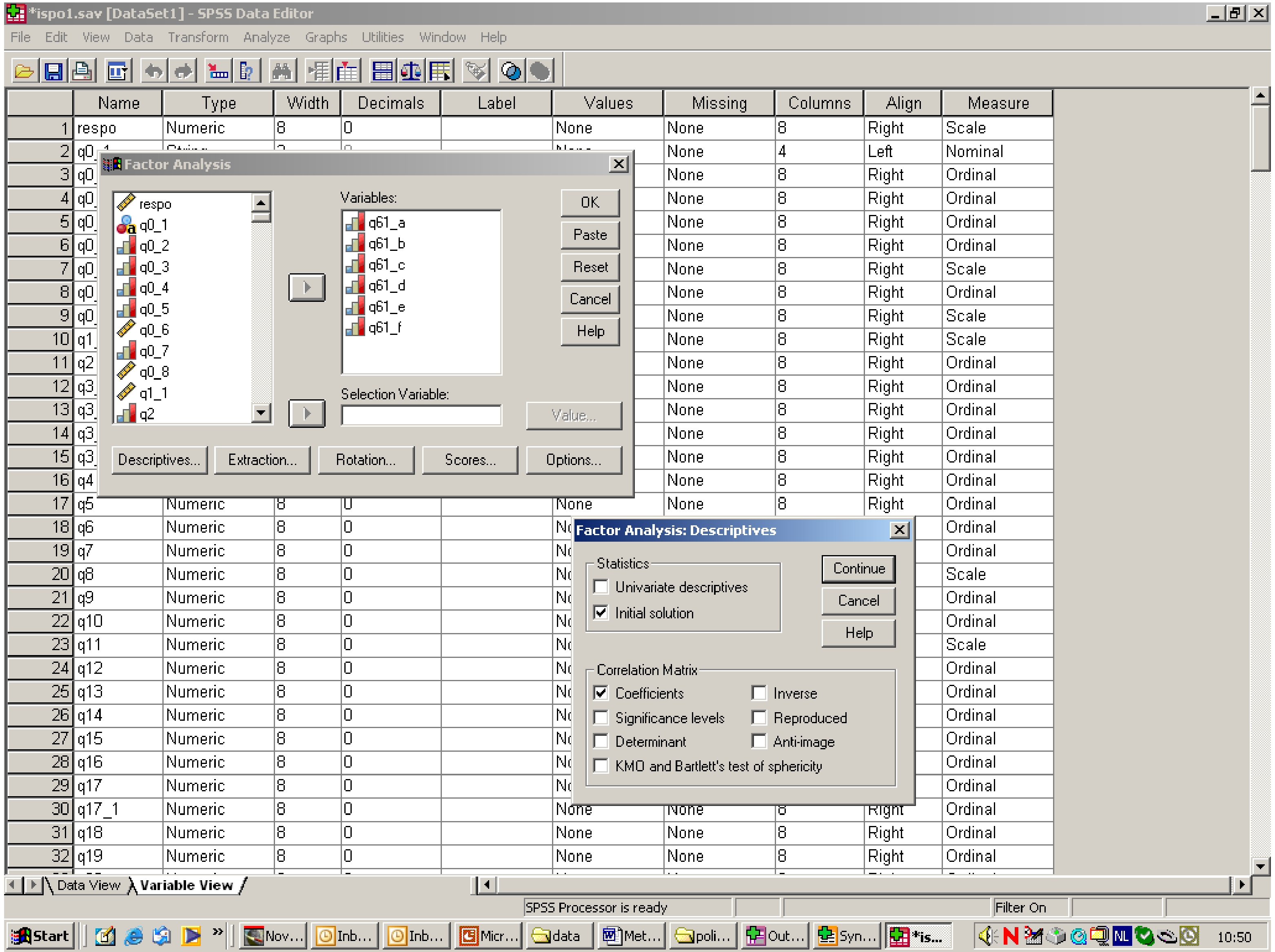
SPSS
Data Reduction: Factor



	Name	Type			Label	Values	Missing	Columns	Align	Measure
1	respo	Numeric				None	None	8	Right	Scale
2	q0_1	String				None	None	4	Left	Nominal
3	q0_2	Numeric				None	None	8	Right	Ordinal
4	q0_3	Numeric				None	None	8	Right	Ordinal
5	q0_4	Numeric				None	None	8	Right	Ordinal
6	q0_5	Numeric				None	None	8	Right	Ordinal
7	q0_6	Numeric				None	None	8	Right	Scale
8	q0_7	Numeric					None	8	Right	Ordinal
9	q0_8	Numeric					None	8	Right	Scale
10	q1_1	Numeric					None	8	Right	Scale
11	q2	Numeric				None	None	8	Right	Ordinal
12	q3_1	Numeric				None	None	8	Right	Ordinal
13	q3_2	Numeric				None	None	8	Right	Ordinal
14	q3_3	Numeric				None	None	8	Right	Ordinal
15	q3_4	Numeric				None	None	8	Right	Ordinal
16	q4	Numeric				None	None	8	Right	Ordinal
17	q5	Numeric				None	None	8	Right	Ordinal
18	q6	Numeric				None	None	8	Right	Ordinal
19	q7	Numeric	8	0		None	None	8	Right	Ordinal
20	q8	Numeric	8	0		None	None	8	Right	Scale
21	q9	Numeric	8	0		None	None	8	Right	Ordinal
22	q10	Numeric	8	0		None	None	8	Right	Ordinal
23	q11	Numeric	8	0		None	None	8	Right	Scale
24	q12	Numeric	8	0		None	None	8	Right	Ordinal
25	q13	Numeric	8	0		None	None	8	Right	Ordinal
26	q14	Numeric	8	0		None	None	8	Right	Ordinal
27	q15	Numeric	8	0		None	None	8	Right	Ordinal
28	q16	Numeric	8	0		None	None	8	Right	Ordinal
29	q17	Numeric	8	0		None	None	8	Right	Ordinal
30	q17_1	Numeric	8	0		None	None	8	Right	Ordinal
31	q18	Numeric	8	0		None	None	8	Right	Ordinal
32	q19	Numeric	8	0		None	None	8	Right	Ordinal

This project has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 900009.

SPSS
Factor analysis
Descriptives



*ispo1.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	respo	Numeric	8	0		None	None	8	Right	Scale
2	q0						None	4	Left	Nominal
3	q0						None	8	Right	Ordinal
4	q0						None	8	Right	Ordinal
5	q0						None	8	Right	Ordinal
6	q0						None	8	Right	Ordinal
7	q0						None	8	Right	Scale
8	q0						None	8	Right	Ordinal
9	q0						None	8	Right	Scale
10	q1						None	8	Right	Scale
11	q2						None	8	Right	Ordinal
12	q3						None	8	Right	Ordinal
13	q3						None	8	Right	Ordinal
14	q3						None	8	Right	Ordinal
15	q3						None	8	Right	Ordinal
16	q4						None	8	Right	Ordinal
17	q5	Numeric	8	0		None	None	8	Right	Ordinal
18	q6	Numeric	8	0		No				Ordinal
19	q7	Numeric	8	0		No				Ordinal
20	q8	Numeric	8	0		No				Scale
21	q9	Numeric	8	0		No				Ordinal
22	q10	Numeric	8	0		No				Ordinal
23	q11	Numeric	8	0		No				Scale
24	q12	Numeric	8	0		No				Ordinal
25	q13	Numeric	8	0		No				Ordinal
26	q14	Numeric	8	0		No				Ordinal
27	q15	Numeric	8	0		No				Ordinal
28	q16	Numeric	8	0		No				Ordinal
29	q17	Numeric	8	0		No				Ordinal
30	q17_1	Numeric	8	0		None	None	8	Right	Ordinal
31	q18	Numeric	8	0		None	None	8	Right	Ordinal
32	q19	Numeric	8	0		None	None	8	Right	Ordinal

Factor Analysis

Source variables: respo, q0_1, q0_2, q0_3, q0_4, q0_5, q0_6, q0_7, q0_8, q1_1, q2

Variables: q61_a, q61_b, q61_c, q61_d, q61_e, q61_f

Selection Variable:

Factor Analysis: Descriptives

Statistics
- ☐ Univariate descriptives
- ☑ Initial solution

Correlation Matrix
- ☑ Coefficients
- ☐ Inverse
- ☐ Significance levels
- ☐ Reproduced
- ☐ Determinant
- ☐ Anti-image
- ☐ KMO and Bartlett's test of sphericity

Political efficacy
Inspired by NES US

Q61.a There's no sense in voting; the *parties* do what they want to do anyway.

No opinion= 5; missing= 1 - teken

Q61.b *Parties* are only interested in my vote, not in my opinion.

No opinion= 6; missing= 2 - teken

Q61.c If people like me let the *politicians* know what we think, then they will take our opinion into account.

No opinion= 52; missing= 1 + teken => spiegelen

Q61.d Most *politicians* promise a lot, but don't do anything.

No opinion= 0; missing= 2 - teken

Q61.e As soon as they are elected, *politicians* think they are better than people like me.

No opinion= 15; missing= 2 - teken

Q61.f Most of our *politicians* are competent people who know what they are doing.

No opinion= 11; missing= 1 + teken => spiegelen

FA – Everything starts with the correlation matrix

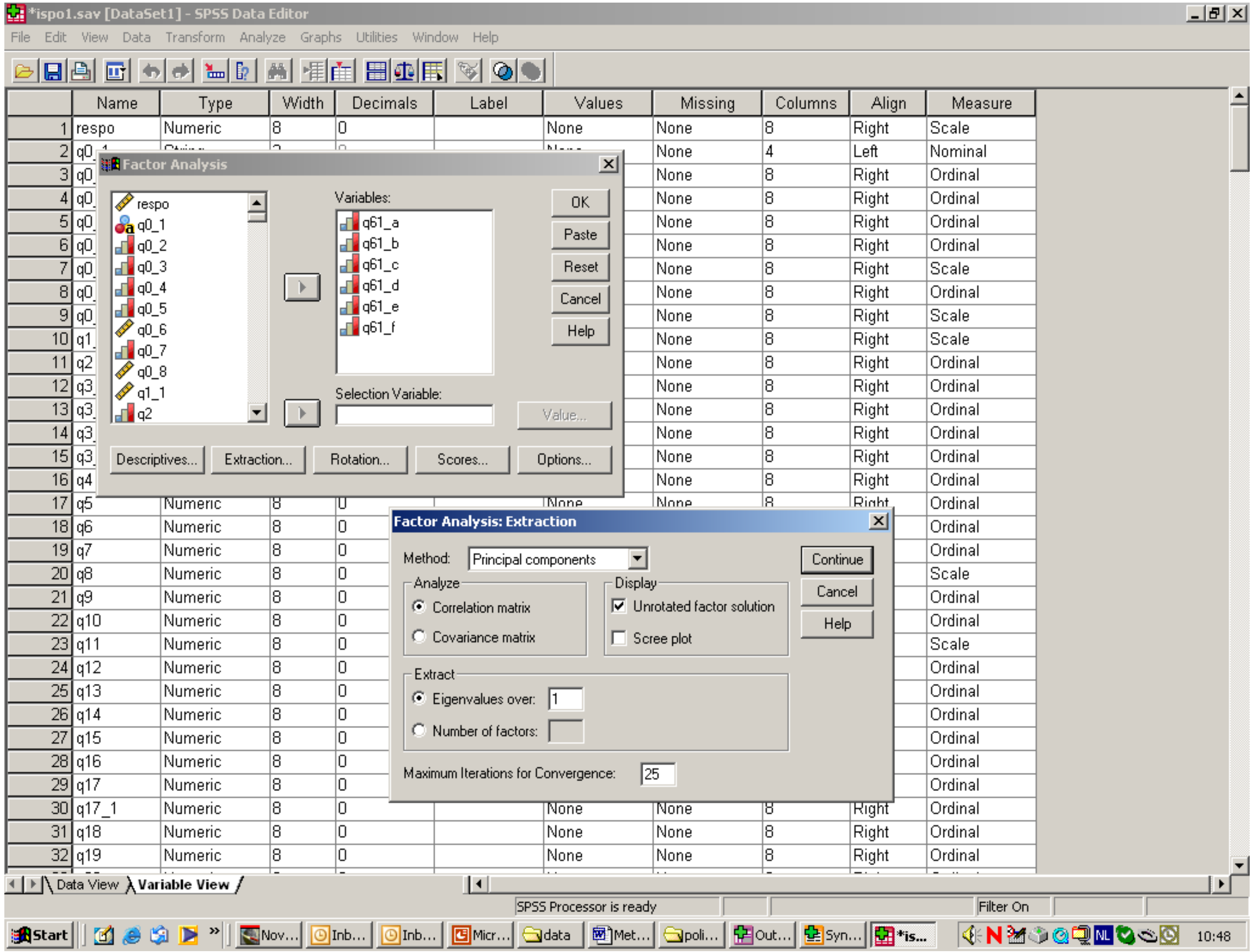
Correlation Matrix

		q61_a	q61_b	q61_c	q61_d	q61_e	q61_f
Correlation	q61_a	1,000	,649	-,325	,483	,502	-,153
	q61_b	,649	1,000	-,398	,520	,550	-,195
	q61_c	-,325	-,398	1,000	-,313	-,332	,182
	q61_d	,483	,520	-,313	1,000	,628	-,233
	q61_e	,502	,550	-,332	,628	1,000	-,238
	q61_f	-,153	-,195	,182	-,233	-,238	1,000

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,817
Bartlett's Test of Sphericity	Approx. Chi-Square	2091,581
	df	15
	Sig.	,000

SPSS
Factor analysis
Descriptives


*ispo1.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
Start
Nov…
Inb…
Inb…
Micr…
data
Met…
poli…
Out…
Syn…
*is…
10:48

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	respo	Numeric	8	0		None	None	8	Right	Scale
2	q0_						None	4	Left	Nominal
3	q0_						None	8	Right	Ordinal
4	q0_						None	8	Right	Ordinal
5	q0_						None	8	Right	Ordinal
6	q0_						None	8	Right	Ordinal
7	q0_						None	8	Right	Scale
8	q0_						None	8	Right	Ordinal
9	q0_						None	8	Right	Scale
10	q1						None	8	Right	Scale
11	q2						None	8	Right	Ordinal
12	q3						None	8	Right	Ordinal
13	q3						None	8	Right	Ordinal
14	q3						None	8	Right	Ordinal
15	q3						None	8	Right	Ordinal
16	q4						None	8	Right	Ordinal
17	q5	Numeric	8	0		None	None	8	Right	Ordinal
18	q6	Numeric	8	0						Ordinal
19	q7	Numeric	8	0						Ordinal
20	q8	Numeric	8	0						Scale
21	q9	Numeric	8	0						Ordinal
22	q10	Numeric	8	0						Ordinal
23	q11	Numeric	8	0						Scale
24	q12	Numeric	8	0						Ordinal
25	q13	Numeric	8	0						Ordinal
26	q14	Numeric	8	0						Ordinal
27	q15	Numeric	8	0						Ordinal
28	q16	Numeric	8	0						Ordinal
29	q17	Numeric	8	0						Ordinal
30	q17_1	Numeric	8	0		None	None	8	Right	Ordinal
31	q18	Numeric	8	0		None	None	8	Right	Ordinal
32	q19	Numeric	8	0		None	None	8	Right	Ordinal

Factor Analysis

respo
q0_1
q0_2
q0_3
q0_4
q0_5
q0_6
q0_7
q0_8
q1_1
q2

Variables:
q61_a
q61_b
q61_c
q61_d
q61_e
q61_f

Selection Variable:

OK | Paste | Reset | Cancel | Help | Value...

Descriptives... | Extraction... | Rotation... | Scores... | Options...

Factor Analysis: Extraction

Method: Principal components

Analyze
● Correlation matrix
○ Covariance matrix

Display
☑ Unrotated factor solution
☐ Scree plot

Extract
● Eigenvalues over: 1
○ Number of factors:

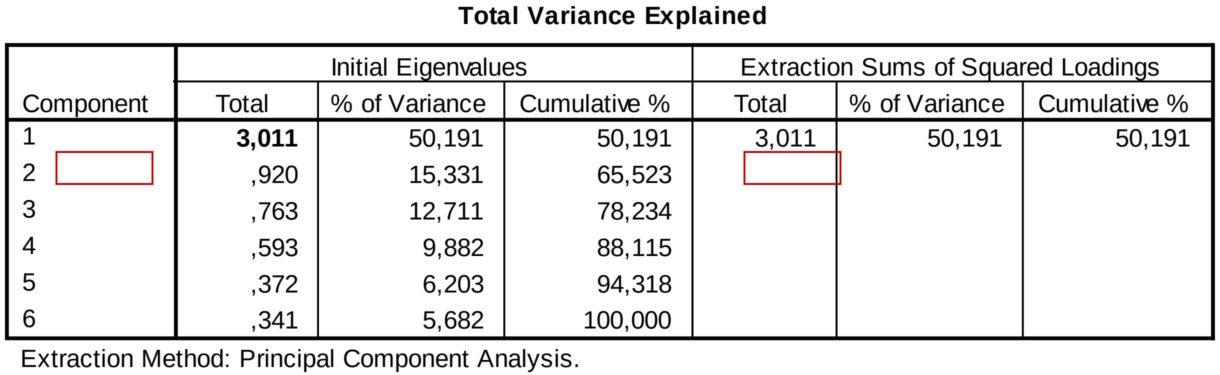
Maximum Iterations for Convergence: 25

Continue | Cancel | Help

SPSS Processor is ready | Filter On

FA – How many factors?

Kaiser’s Criterion: factors with an eigenvalue higher than 1 or an explained variance of at least 60%…



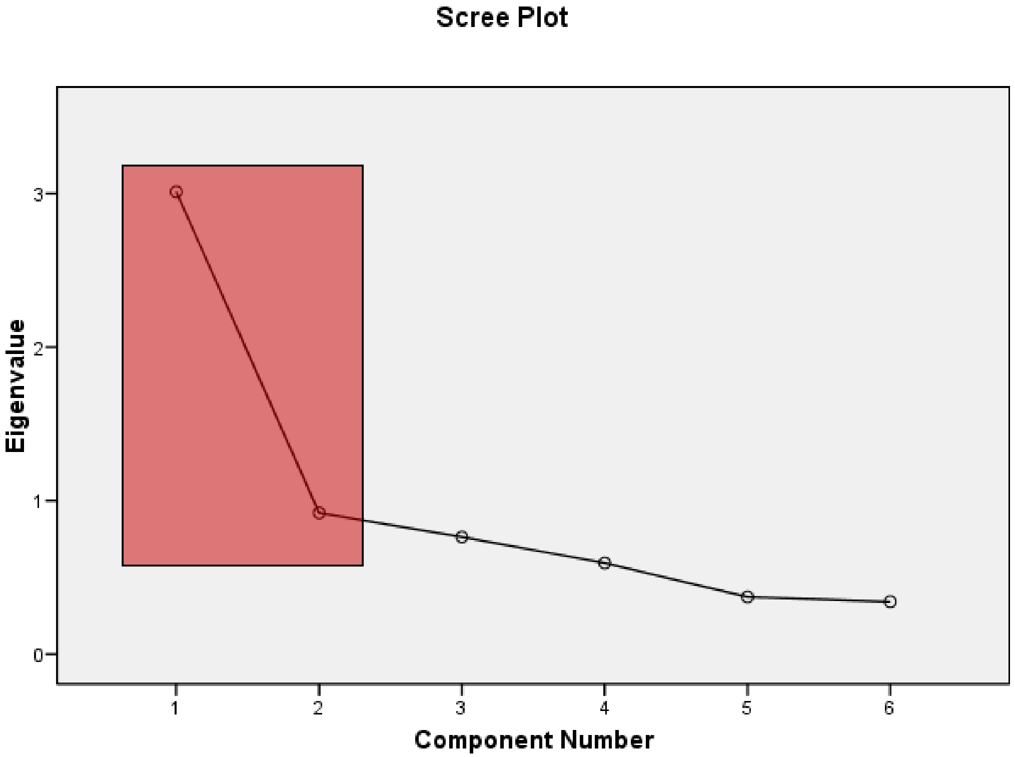
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	**3,011**	50,191	50,191	3,011	50,191	50,191
2	,920	15,331	65,523			
3	,763	12,711	78,234			
4	,593	9,882	88,115			
5	,372	6,203	94,318			
6	,341	5,682	100,000			

Extraction Method: Principal Component Analysis.

FA – How many factors?

Cattell’s Criterion: looking for the elbow in a scree plot of the eigenvalues

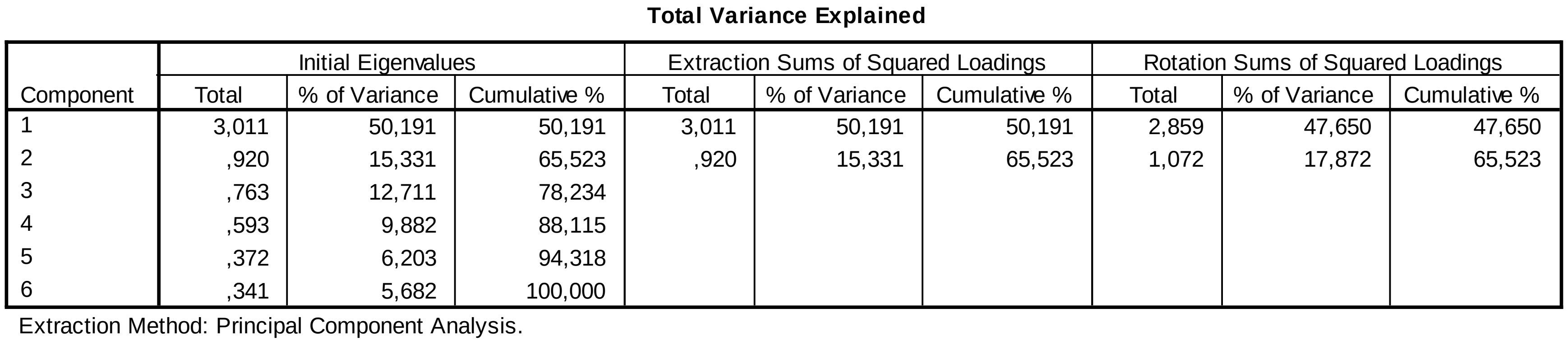


FA – How many factors ?

Criterion of Theo Ry:

Do you see a valid theoretical explanation for a certain dimensionalization?

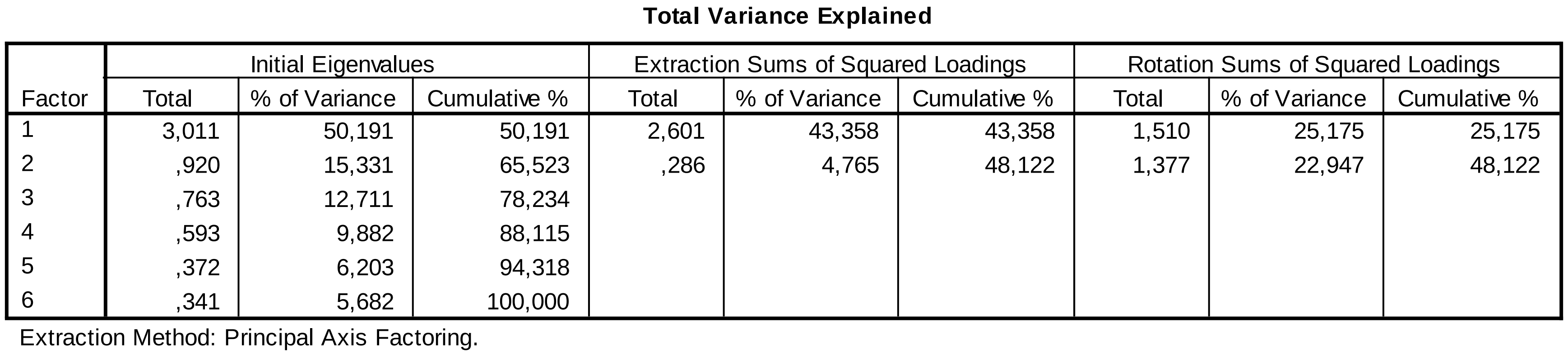
FA – How many factors?



Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,011	50,191	50,191	3,011	50,191	50,191	2,859	47,650	47,650
2	,920	15,331	65,523	,920	15,331	65,523	1,072	17,872	65,523
3	,763	12,711	78,234						
4	,593	9,882	88,115						
5	,372	6,203	94,318						
6	,341	5,682	100,000						

Extraction Method: Principal Component Analysis.

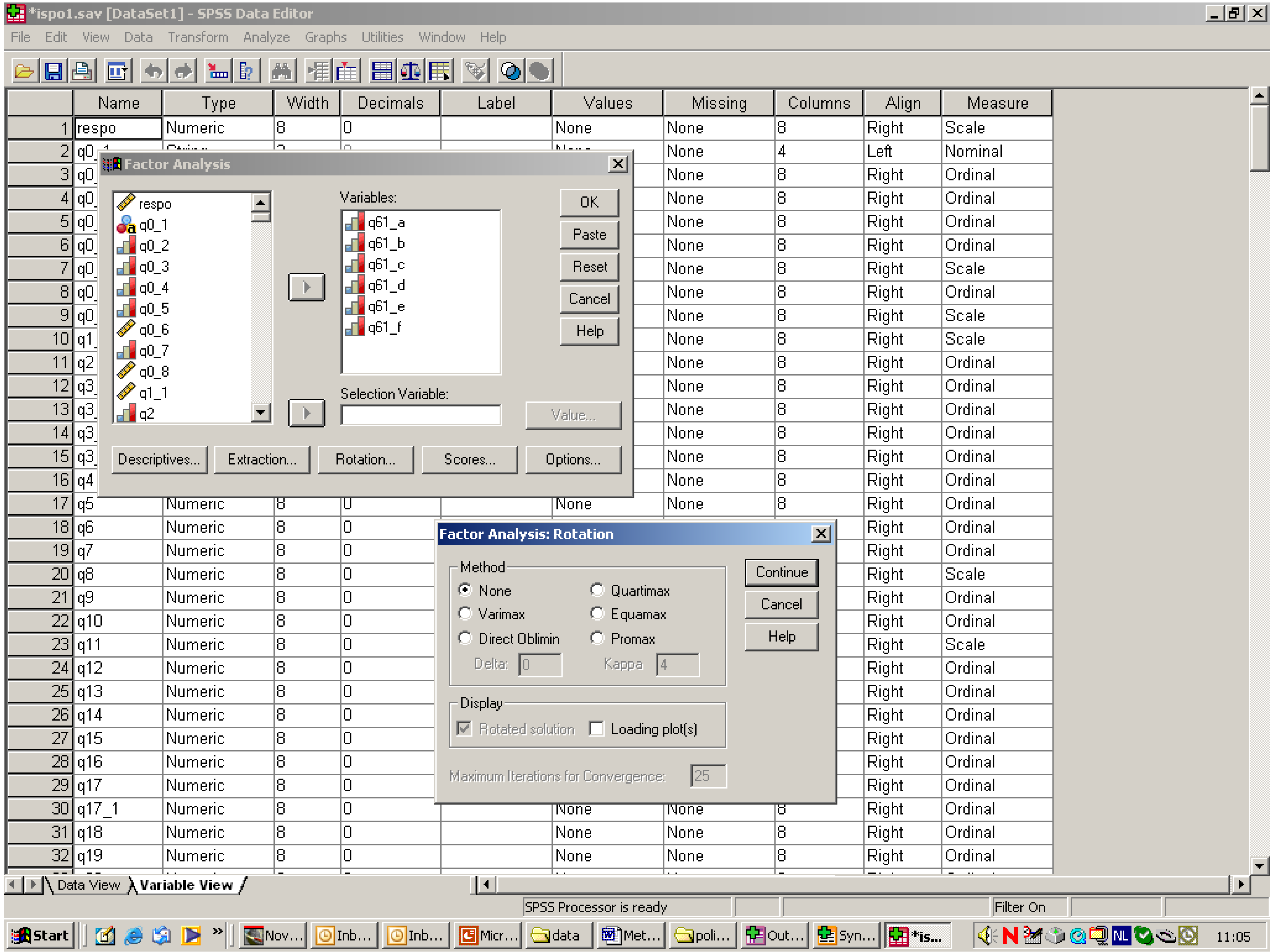


Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,011	50,191	50,191	2,601	43,358	43,358	1,510	25,175	25,175
2	,920	15,331	65,523	,286	4,765	48,122	1,377	22,947	48,122
3	,763	12,711	78,234						
4	,593	9,882	88,115						
5	,372	6,203	94,318						
6	,341	5,682	100,000						

Extraction Method: Principal Axis Factoring.

SPSS
Factor analysis
Rotation



*ispo1.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	respo	Numeric	8	0		None	None	8	Right	Scale
2	q0_1						None	4	Left	Nominal
3	q0_						None	8	Right	Ordinal
4	q0_						None	8	Right	Ordinal
5	q0_						None	8	Right	Ordinal
6	q0_						None	8	Right	Ordinal
7	q0_						None	8	Right	Scale
8	q0_						None	8	Right	Ordinal
9	q0_						None	8	Right	Scale
10	q1_						None	8	Right	Scale
11	q2						None	8	Right	Ordinal
12	q3_						None	8	Right	Ordinal
13	q3_						None	8	Right	Ordinal
14	q3_						None	8	Right	Ordinal
15	q3_						None	8	Right	Ordinal
16	q4						None	8	Right	Ordinal
17	q5	Numeric	8	0		None	None	8	Right	Ordinal
18	q6	Numeric	8	0					Right	Ordinal
19	q7	Numeric	8	0					Right	Ordinal
20	q8	Numeric	8	0					Right	Scale
21	q9	Numeric	8	0					Right	Ordinal
22	q10	Numeric	8	0					Right	Ordinal
23	q11	Numeric	8	0					Right	Scale
24	q12	Numeric	8	0					Right	Ordinal
25	q13	Numeric	8	0					Right	Ordinal
26	q14	Numeric	8	0					Right	Ordinal
27	q15	Numeric	8	0					Right	Ordinal
28	q16	Numeric	8	0					Right	Ordinal
29	q17	Numeric	8	0					Right	Ordinal
30	q17_1	Numeric	8	0		None	None	8	Right	Ordinal
31	q18	Numeric	8	0		None	None	8	Right	Ordinal
32	q19	Numeric	8	0		None	None	8	Right	Ordinal

Factor Analysis

respo
q0_1
q0_2
q0_3
q0_4
q0_5
q0_6
q0_7
q0_8
q1_1
q2

Variables:
q61_a
q61_b
q61_c
q61_d
q61_e
q61_f

Selection Variable:

Factor Analysis: Rotation

Method
● None ○ Quartimax
○ Varimax ○ Equamax
○ Direct Oblimin ○ Promax
Delta: 0 Kappa 4

Display
☑ Rotated solution □ Loading plot(s)

Maximum Iterations for Convergence: 25

Looking for a 'simple structure'

Via rotation:

Orthogonal

Varimax

Meaningful factor loadings: rule of thumb > 0,50

Political efficacy
Inspired by NES US

Q61.a There's no sense in voting; the *parties* do what they want to do anyway.

No opinion= 5; missing= 1 - teken

Q61.b *Parties* are only interested in my vote, not in my opinion.
No opinion= 6; missing= 2 - teken

Q61.c If people like me let the *politicians* know what we think, then they will take our opinion into account.

No opinion= 52; missing= 1 + teken => spiegelen

Q61.d Most *politicians* promise a lot, but don't do anything.

No opinion= 0; missing= 2 - teken

Q61.e As soon as they are elected, *politicians* think they are better than people like me.

No opinion= 15; missing= 2 - teken

Q61.f Most of our *politicians* are competent people who know what they are doing.
No opinion= 11; missing= 1 + teken => spiegelen

PCA versus PFA

Communalities

	Initial	Extraction
q61_a	1,000	,664
q61_b	1,000	,708
q61_c	1,000	,339
q61_d	1,000	,614
q61_e	1,000	,646
q61_f	1,000	,960

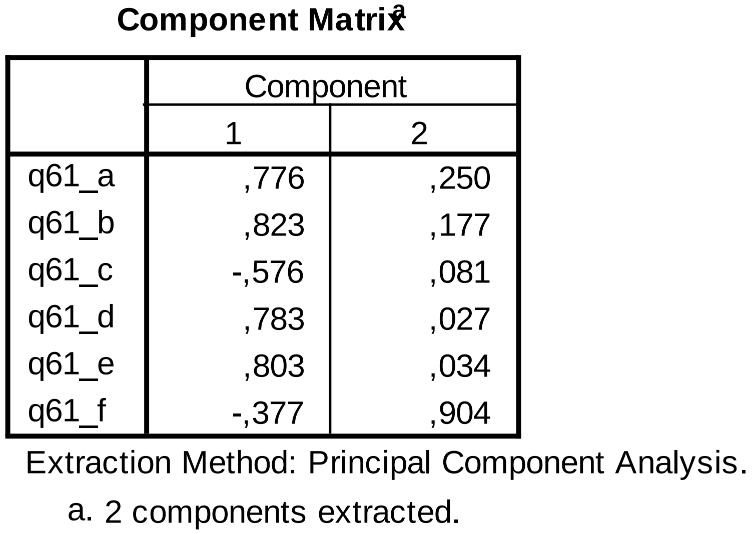
Extraction Method: Principal Component Analysis.

Communalities

	Initial	Extraction
q61_a	,463	,537
q61_b	,522	,798
q61_c	,190	,211
q61_d	,455	,606
q61_e	,482	,643
q61_f	,078	,093

Extraction Method: Principal Axis Factoring.

FA - ‘Simple structure’

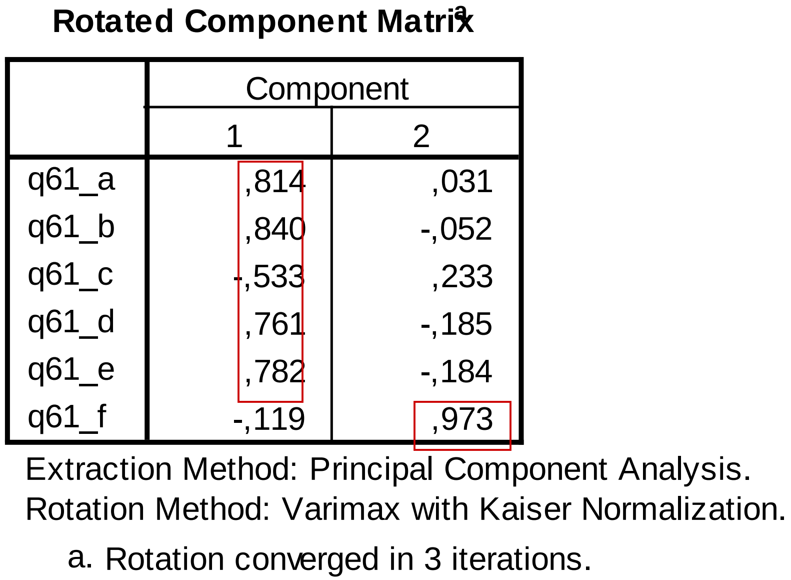


Component Matrix[a]

	Component	
	1	2
q61_a	,776	,250
q61_b	,823	,177
q61_c	-,576	,081
q61_d	,783	,027
q61_e	,803	,034
q61_f	-,377	,904

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

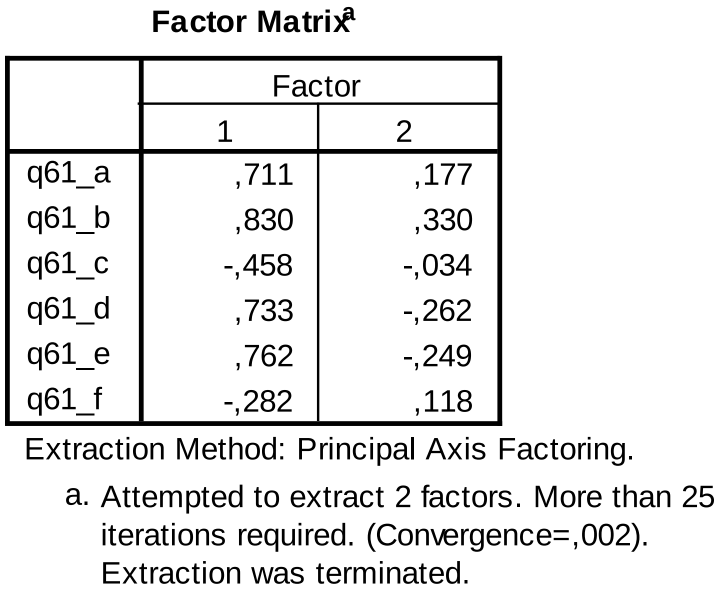


Rotated Component Matrix[a]

	Component	
	1	2
q61_a	,814	,031
q61_b	,840	-,052
q61_c	-,533	,233
q61_d	,761	-,185
q61_e	,782	-,184
q61_f	-,119	,973

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

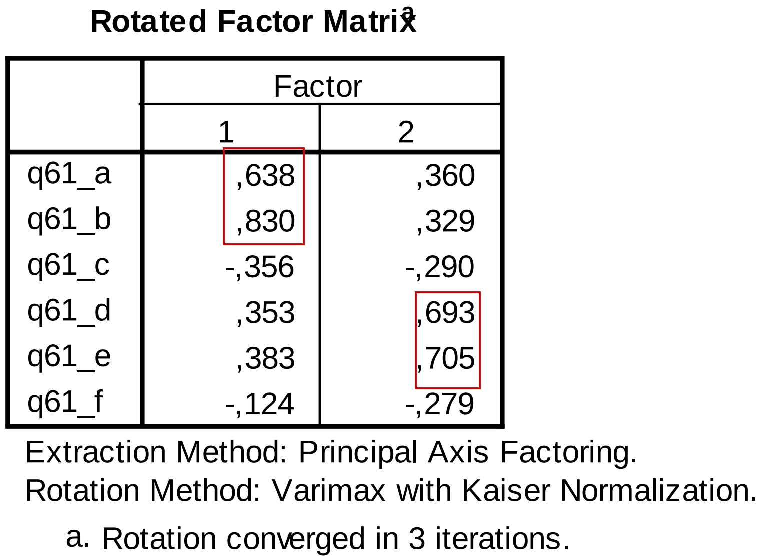


Factor Matrix[a]

	Factor	
	1	2
q61_a	,711	,177
q61_b	,830	,330
q61_c	-,458	-,034
q61_d	,733	-,262
q61_e	,762	-,249
q61_f	-,282	,118

Extraction Method: Principal Axis Factoring.

a. Attempted to extract 2 factors. More than 25 iterations required. (Convergence=,002). Extraction was terminated.



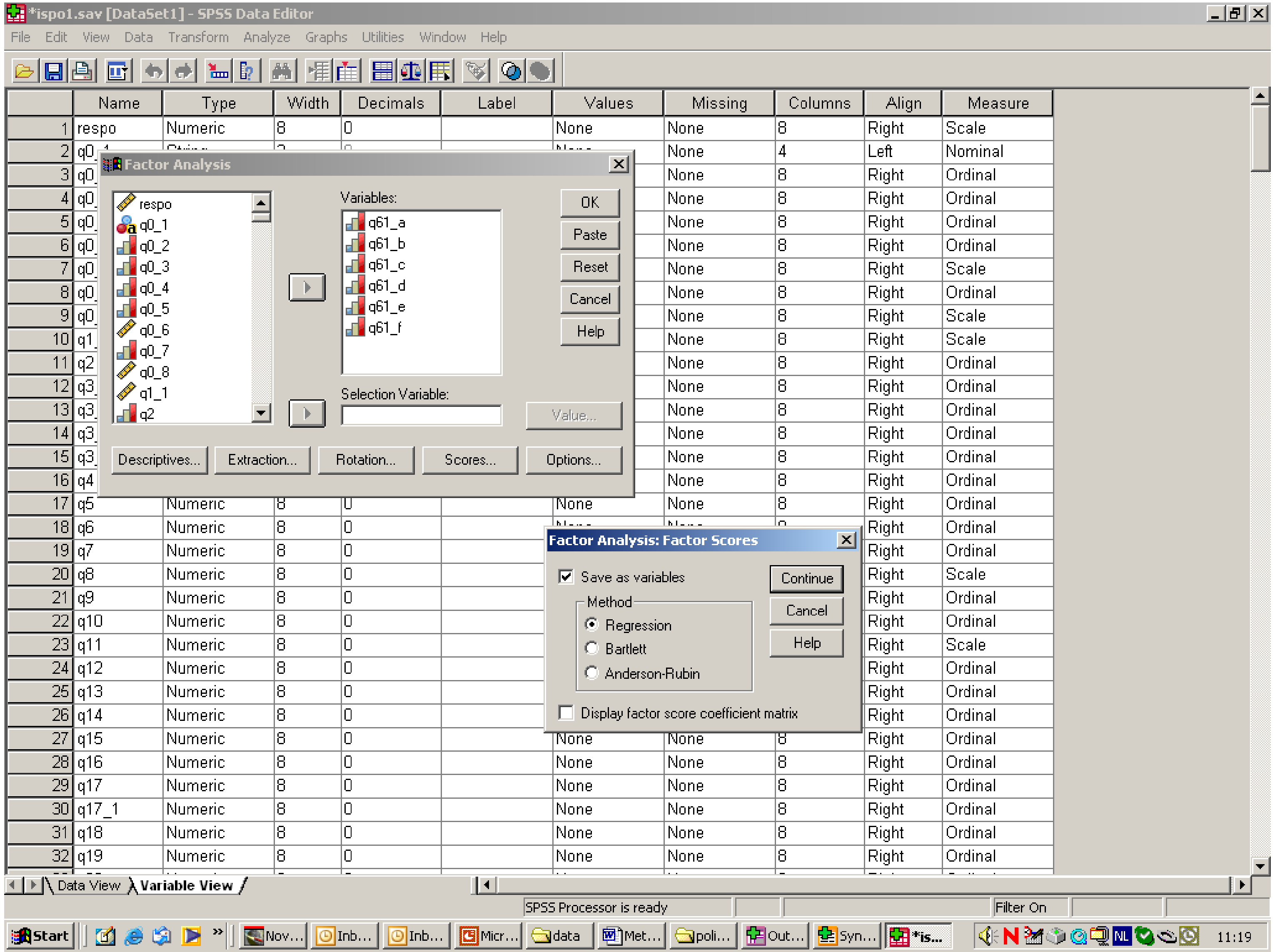
Rotated Factor Matrix[a]

	Factor	
	1	2
q61_a	,638	,360
q61_b	,830	,329
q61_c	-,356	-,290
q61_d	,353	,693
q61_e	,383	,705
q61_f	-,124	-,279

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

SPSS
Factor analysis
Rotation



*ispo1.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready
Filter On
Start
11:19

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	respo	Numeric	8	0		None	None	8	Right	Scale
2	q0_1						None	4	Left	Nominal
3	q0						None	8	Right	Ordinal
4	q0						None	8	Right	Ordinal
5	q0						None	8	Right	Ordinal
6	q0						None	8	Right	Ordinal
7	q0						None	8	Right	Scale
8	q0						None	8	Right	Ordinal
9	q0						None	8	Right	Scale
10	q1						None	8	Right	Scale
11	q2						None	8	Right	Ordinal
12	q3						None	8	Right	Ordinal
13	q3						None	8	Right	Ordinal
14	q3						None	8	Right	Ordinal
15	q3						None	8	Right	Ordinal
16	q4						None	8	Right	Ordinal
17	q5	Numeric	8	0		None	None	8	Right	Ordinal
18	q6	Numeric	8	0					Right	Ordinal
19	q7	Numeric	8	0					Right	Ordinal
20	q8	Numeric	8	0					Right	Scale
21	q9	Numeric	8	0					Right	Ordinal
22	q10	Numeric	8	0					Right	Ordinal
23	q11	Numeric	8	0					Right	Scale
24	q12	Numeric	8	0					Right	Ordinal
25	q13	Numeric	8	0					Right	Ordinal
26	q14	Numeric	8	0					Right	Ordinal
27	q15	Numeric	8	0		None	None	8	Right	Ordinal
28	q16	Numeric	8	0		None	None	8	Right	Ordinal
29	q17	Numeric	8	0		None	None	8	Right	Ordinal
30	q17_1	Numeric	8	0		None	None	8	Right	Ordinal
31	q18	Numeric	8	0		None	None	8	Right	Ordinal
32	q19	Numeric	8	0		None	None	8	Right	Ordinal

Factor Analysis

respo
q0_1
q0_2
q0_3
q0_4
q0_5
q0_6
q0_7
q0_8
q1_1
q2

Variables:
q61_a
q61_b
q61_c
q61_d
q61_e
q61_f

Selection Variable:

OK | Paste | Reset | Cancel | Help | Value...

Descriptives... | Extraction... | Rotation... | Scores... | Options...

Factor Analysis: Factor Scores

☑ Save as variables

Method
- ● Regression
- ○ Bartlett
- ○ Anderson-Rubin

□ Display factor score coefficient matrix

Continue | Cancel | Help

Saving and explaining factorscores

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,360[a]	,129	,128	,85855727

a. Predictors: (Constant), age_vla, autor

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1,290	,108		11,908	,000
	autor	-,182	,015	-,349	-12,558	,000
	age_vla	-,133	,061	-,061	-2,195	,028

a. Dependent Variable: REGR factor score 1 for analysis 1

Reliability analysis

In order to test the internal consistency of indicators as measures of a unidimensional latent construct

This project has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 900009.

Reliable Indicators

$$x_i = \tau_i + e_i$$

Test-retest reliability

-> correlations over time r(x_{t1}, x_{t2}) of r(x_{t1}, y_{t2})

BUT trade-off reminder – real change

Internal consistency

-> split-half r($\sum x_{helft1}$, $\sum x_{helft2}$)

BUT many possible partitions

-> Cronbach’s alpha: mean correlation of all possible partitions

Cronbach’s alpha (1)

Alpha= proportion common variance

Covariances = common variance

$$\sigma_{12} = \mathrm{cov}\,ar\,(x_1 x_2) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)$$

Individual variance = unique variance

$$\sigma_1^2 = \mathrm{var}\,(x_1) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2 = \frac{\sum_{i=1}^{n} x_{i1}^2}{n-1} - \frac{n\bar{x}^2}{n-1}$$

Cronbach’s alpha (2)

Variance of scale scores = sum scores

$$\mathrm{var}\,(x_1 + x_2) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i1} + x_{i2} - \bar{x}_1 - \bar{x}_2)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[(x_{i1} - \bar{x}_1) + (x_{i2} - \bar{x}_2)\right]^2 =$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2 + \frac{1}{n-1}\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2 + \frac{2}{n-1}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) =$$

$$\underbrace{\sigma_1^2 + \sigma_2^2}_{\text{Unique variance}} + \underbrace{2\sigma_{12}}_{\text{Common variance}}$$

Cronbach's alpha (3)

Variance of scale scores = sum scores

-> Logic for 4 items

$$\operatorname{var} S = \operatorname{var}(x_1 + x_2 + x_3 + x_4) =$$

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 + \sigma_{12} + \sigma_{21} + \sigma_{13} + \sigma_{31} + \sigma_{14} + \sigma_{41} + \sigma_{23} + \sigma_{32} + \sigma_{24} + \sigma_{42} + \sigma_{34} + \sigma_{43}$$

Var S_i= $$\sum_{i=1}^{n} \sigma_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} \sigma_{ij}$$

= unique variance + common variance

Cronbach’s alpha (4)

$$\alpha = \frac{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\sigma_{ij}}{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\sigma_{ij} + \sum_{i=1}^{n}\sigma_i^2} = \frac{ESS}{TSS}$$

Alpha is comparable with R^2

Problem: more items => alpha higher

Cronbach’s alpha (5)

$$\alpha_{adj} = \frac{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\sigma_{ij} \bigg/ n^2-n}{\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\sigma_{ij}+\sum_{i=1}^{n}\sigma_i^2\right) \bigg/ n^2} = \frac{n}{(n-1)}\cdot\alpha \qquad \mathit{because}\; n^2-n=n\cdot(n-1)$$

n.n=n^2 elements in covariance matrix

with n diagonal elements

Adjusted alpha comparable with adjusted R^2

Political efficacy
Inspired by NES US

Q61.a There's no sense in voting; the *parties* do what they want to do anyway.

No opinion= 5; missing= 1 - teken

Q61.b *Parties* are only interested in my vote, not in my opinion.

No opinion= 6; missing= 2 - teken

Q61.c If people like me let the *politicians* know what we think, then they will take our opinion into account.

No opinion= 52; missing= 1 + teken => spiegelen

Q61.d Most *politicians* promise a lot, but don't do anything.

No opinion= 0; missing= 2 - teken

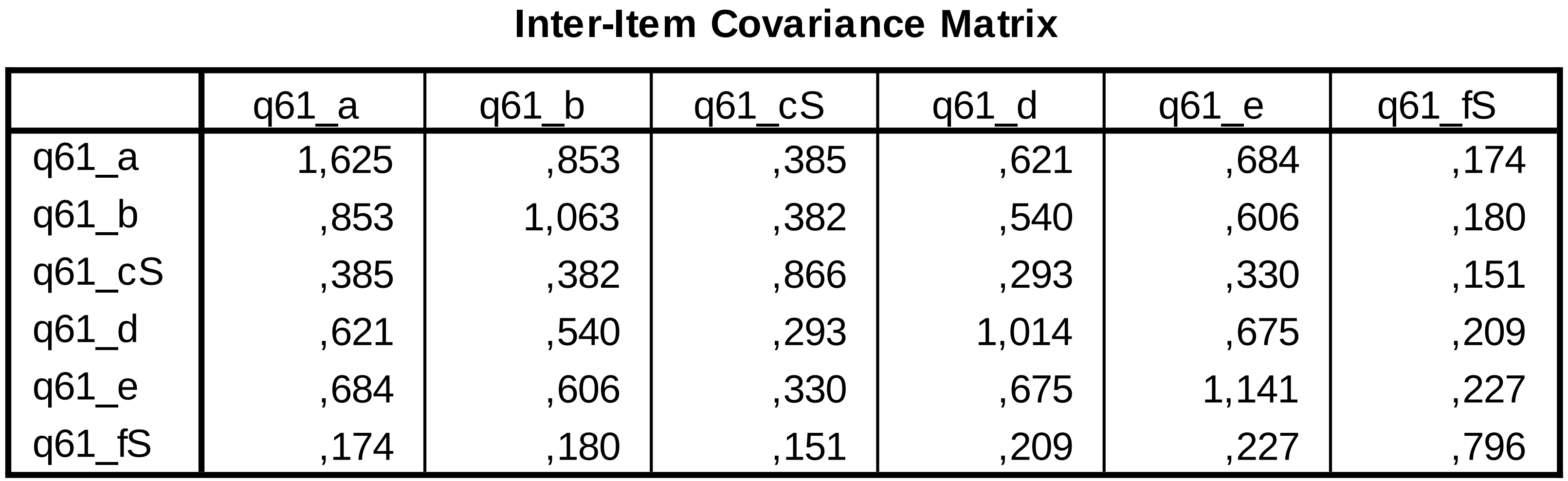
Q61.e As soon as they are elected, *politicians* think they are better than people like me.

No opinion= 15; missing= 2 - teken

Q61.f Most of our *politicians* are competent people who know what they are doing.

No opinion= 11; missing= 1 + teken => spiegelen

Political efficacy – Flanders Covariance matrix



Inter-Item Covariance Matrix

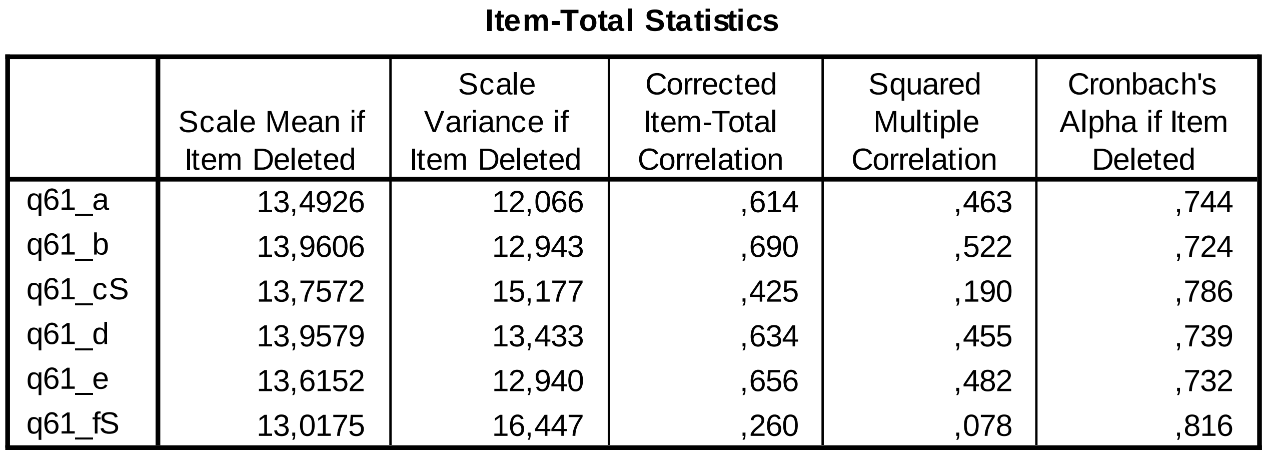
	q61_a	q61_b	q61_cS	q61_d	q61_e	q61_fS
q61_a	1,625	,853	,385	,621	,684	,174
q61_b	,853	1,063	,382	,540	,606	,180
q61_cS	,385	,382	,866	,293	,330	,151
q61_d	,621	,540	,293	1,014	,675	,209
q61_e	,684	,606	,330	,675	1,141	,227
q61_fS	,174	,180	,151	,209	,227	,796

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\sigma_{ij} = (2.0{,}853) + (2.0{,}385) + \ldots + (2.0{,}227) = 12{,}620$$

$$\sum_{i=1}^{n}\sigma_i^2 = 1{,}652 + 1{,}063 + 0{,}866 + 1{,}014 + 1{,}141 + 0{,}796 = 6{,}505$$

Political efficacy – Flanders (Belgium) Cronbach's alpha (6 items)

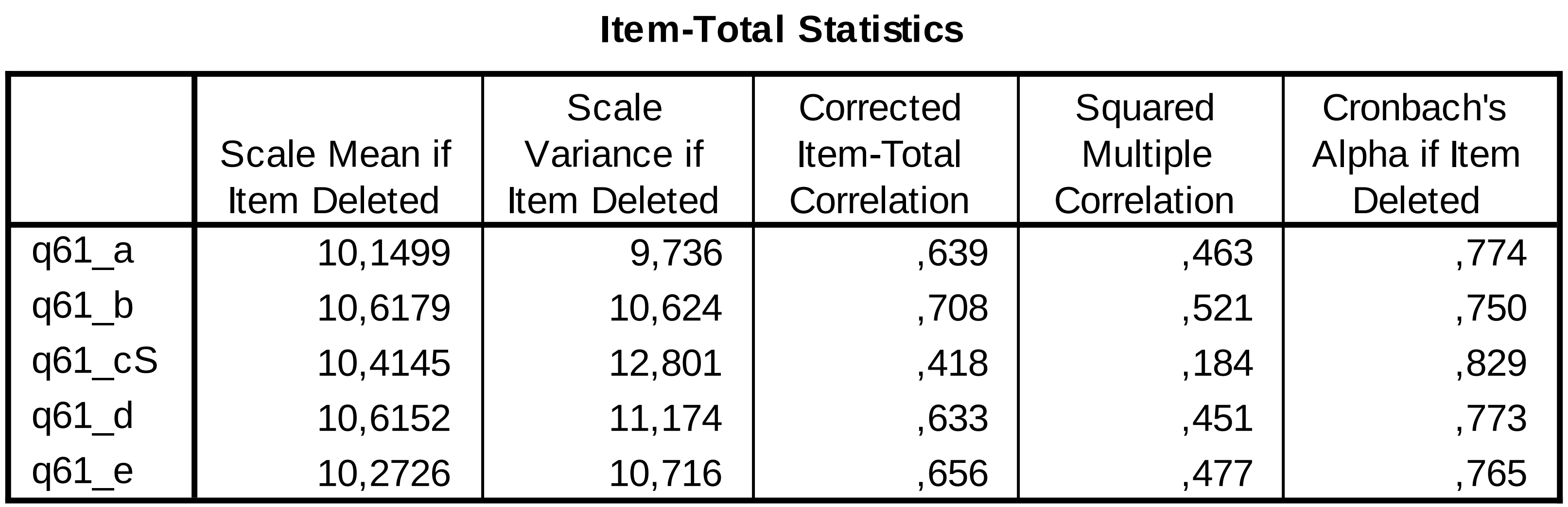
$$\alpha_{adj} = \frac{n}{(n-1)} \cdot \frac{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\sigma_{ij}}{\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\sigma_{ij} + \sum_{i=1}^{n}\sigma_i^2} = \frac{6}{(6-1)} \cdot \frac{12{,}620}{(12{,}620+6{,}505)} = 0{,}792$$



Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
q61_a	13,4926	12,066	,614	,463	,744
q61_b	13,9606	12,943	,690	,522	,724
q61_cS	13,7572	15,177	,425	,190	,786
q61_d	13,9579	13,433	,634	,455	,739
q61_e	13,6152	12,940	,656	,482	,732
q61_fS	13,0175	16,447	,260	,078	,816

Political efficacy – Flanders (Belgium)
Cronbach's alpha (6 items)



Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
q61_a	10,1499	9,736	,639	,463	,774
q61_b	10,6179	10,624	,708	,521	,750
q61_cS	10,4145	12,801	,418	,184	,829
q61_d	10,6152	11,174	,633	,451	,773
q61_e	10,2726	10,716	,656	,477	,765

Cluster analysis

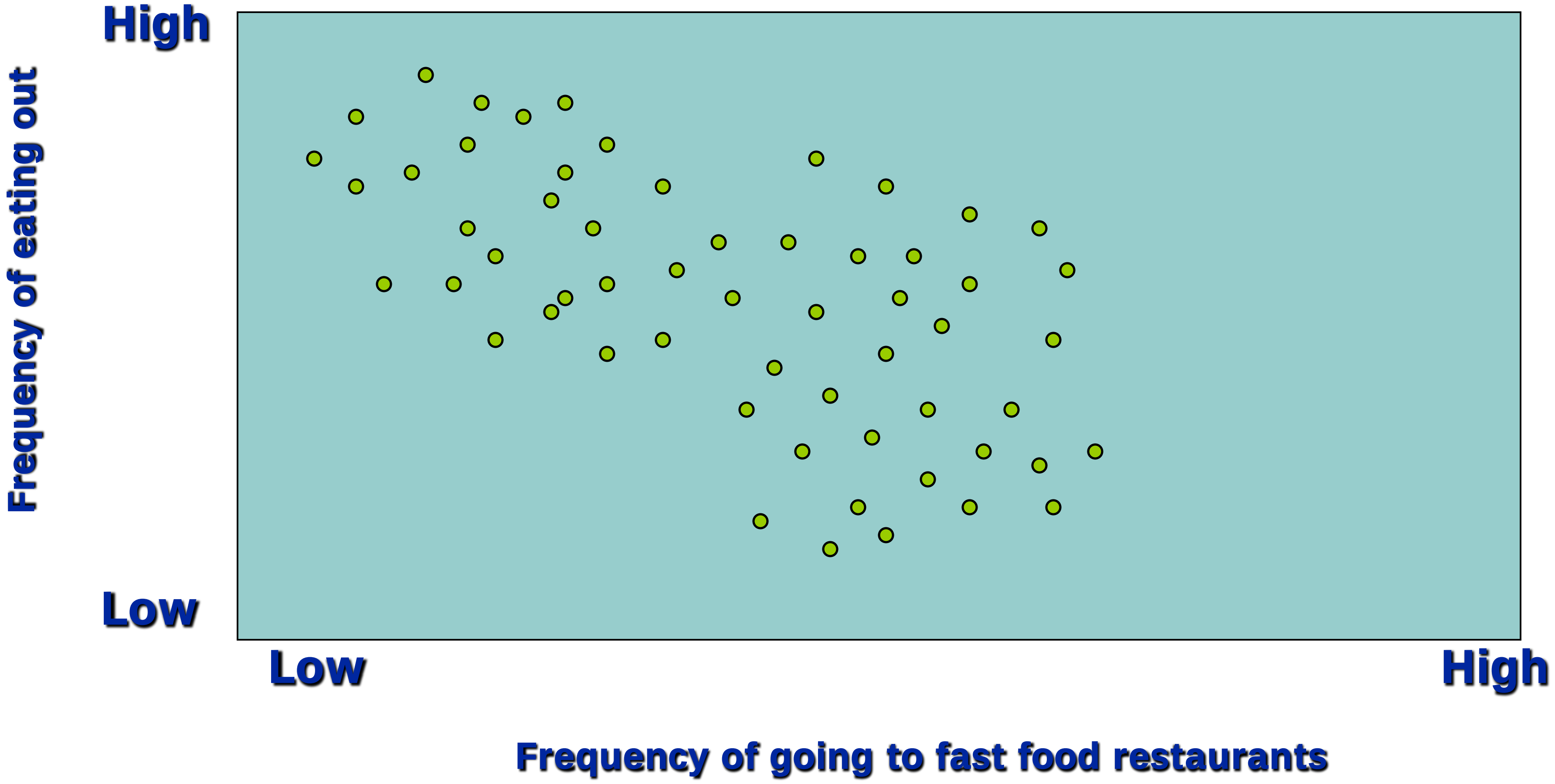
In order to construct groups of respondents that are internally homogeneous and externally heterogeneous based on a set of quantitative indicators

Cluster analysis characteristics

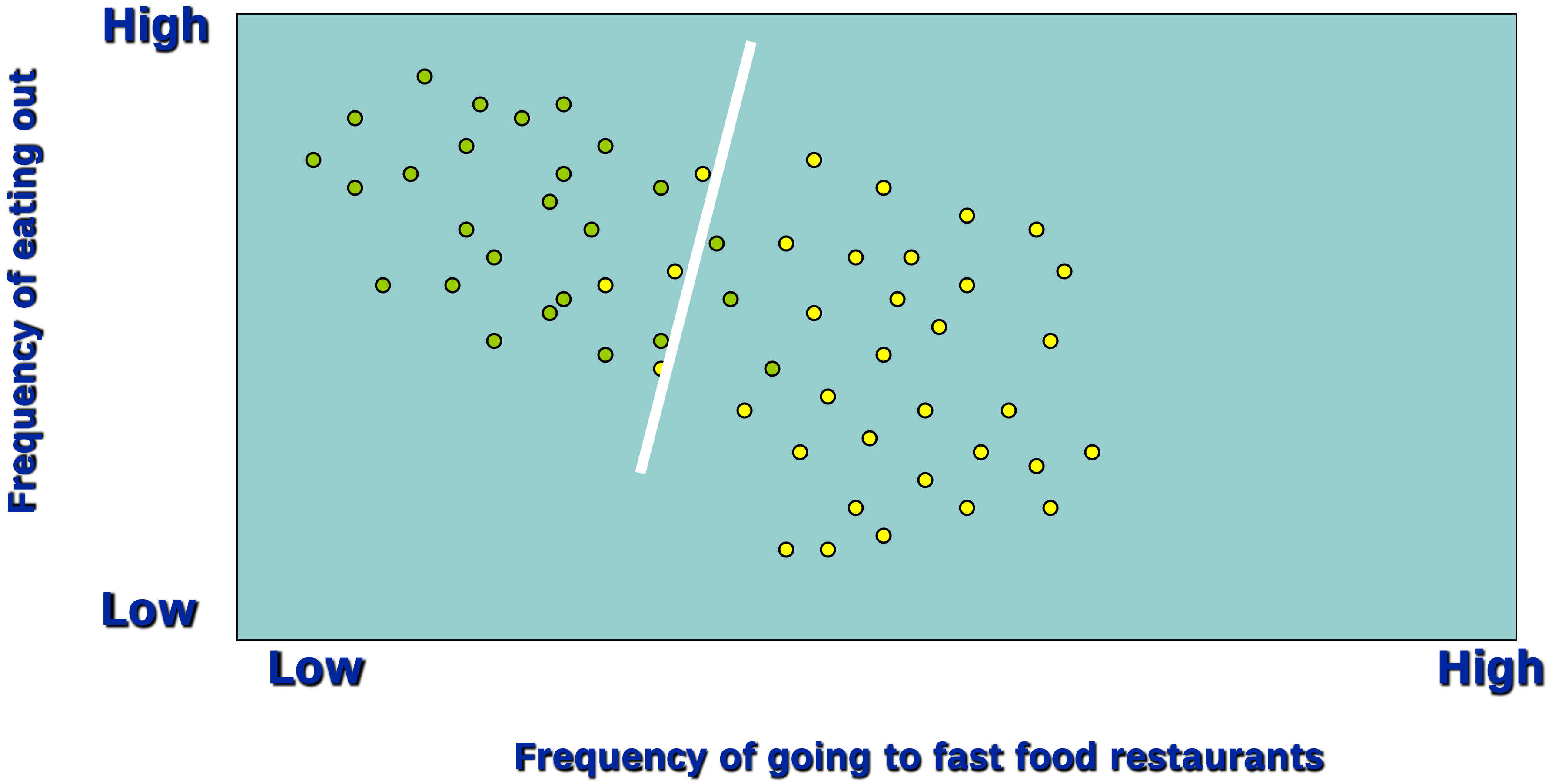
- Data reductive technique
- Symmetric technique
- Explorative, inductive, descriptive
  - Garbage in, Garbage out
- Q-technique (cases) versus R-technique (variables)
  - Clusters: focus on the rows of a data-matrix
- Clusters versus factoren

No unique solution:

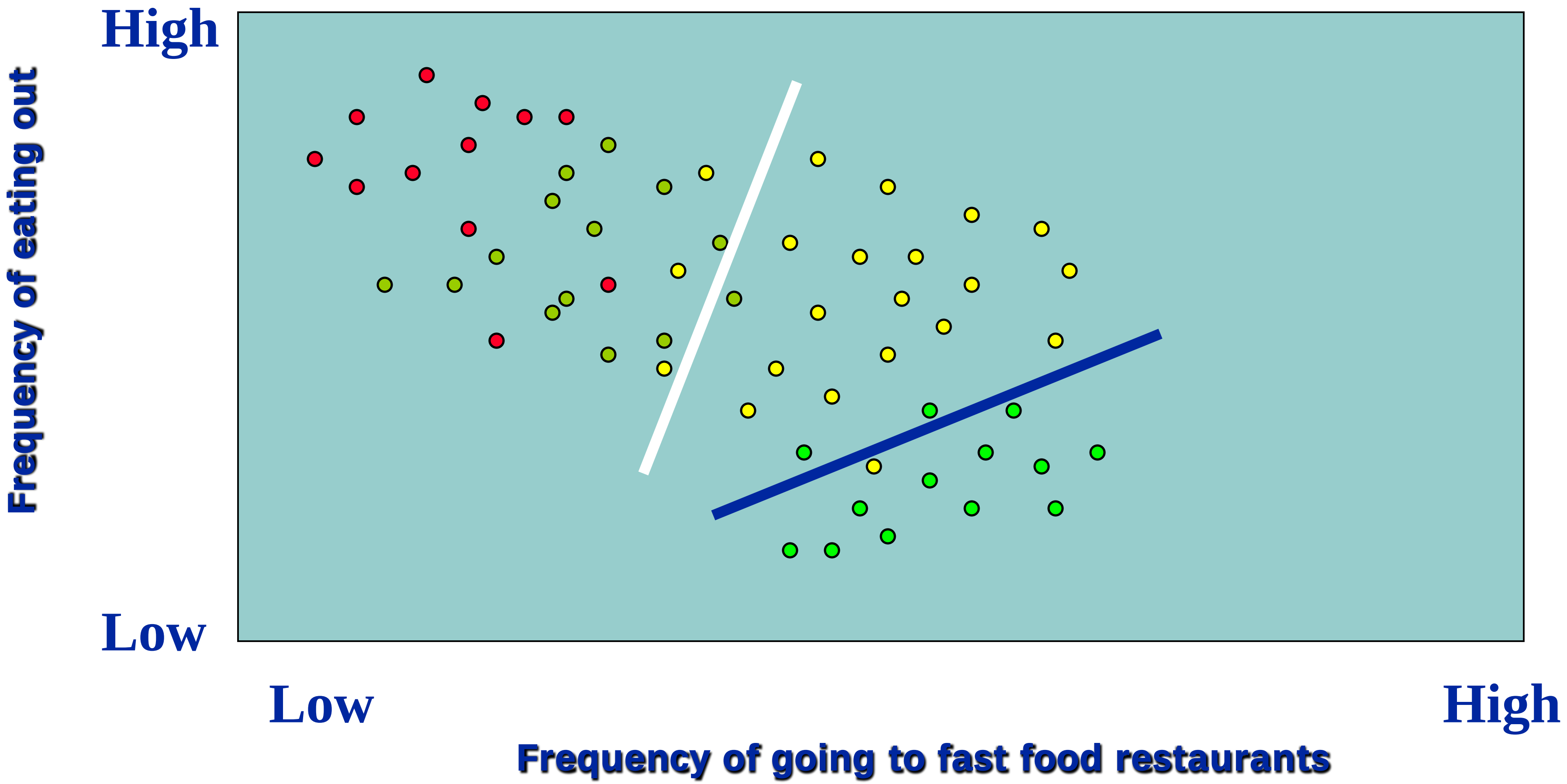
!!! cluster analysis always generates clusters



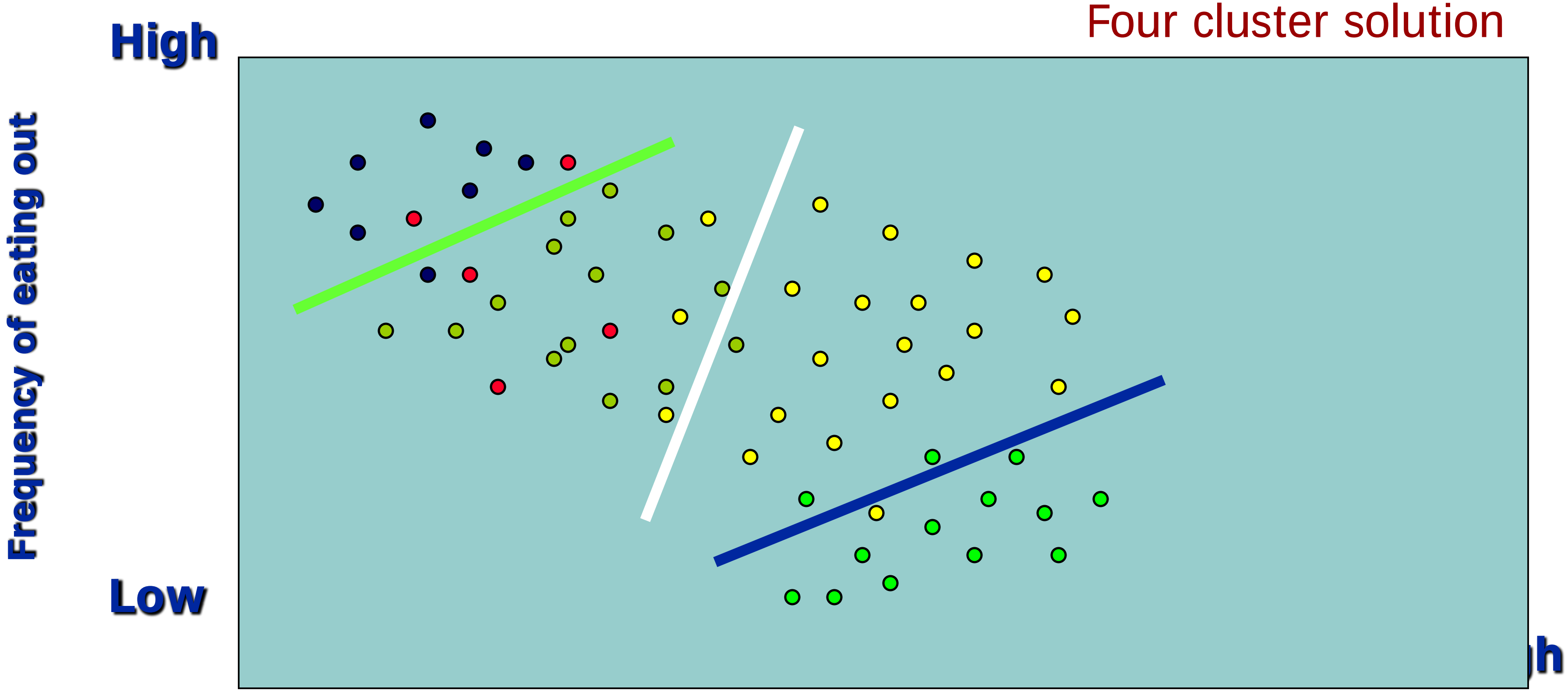
Two cluster solution



Three cluster solution



Four cluster solution



Frequency of going to fast food restaurants

Clusteranalyse: Objectieven

- Opstellen van een classificatie van cases
  - cf. taxonomie/typologie *(analogie met planten, dieren, psychiatrische taxonomy)*
- Reduceren van de complexiteit tussen de cases
- (vaak) Tussenstap in de globale analyse

van fundamenteel belang...

- Relevante variabelen
  - variabele impliceert variatie
  - analyseniveau (relatief versus absoluut)

	V1	V2	V3
case1			
case2			
case3			

- Geschikte onderzoekselementen
  - hiaten of uitschieters
    - boxdiagram of multivariate maatstaf

3 important questions….

1. What kind of measure do we use to assess the likeness or similarity of cases ? Do we need a standardized measure?

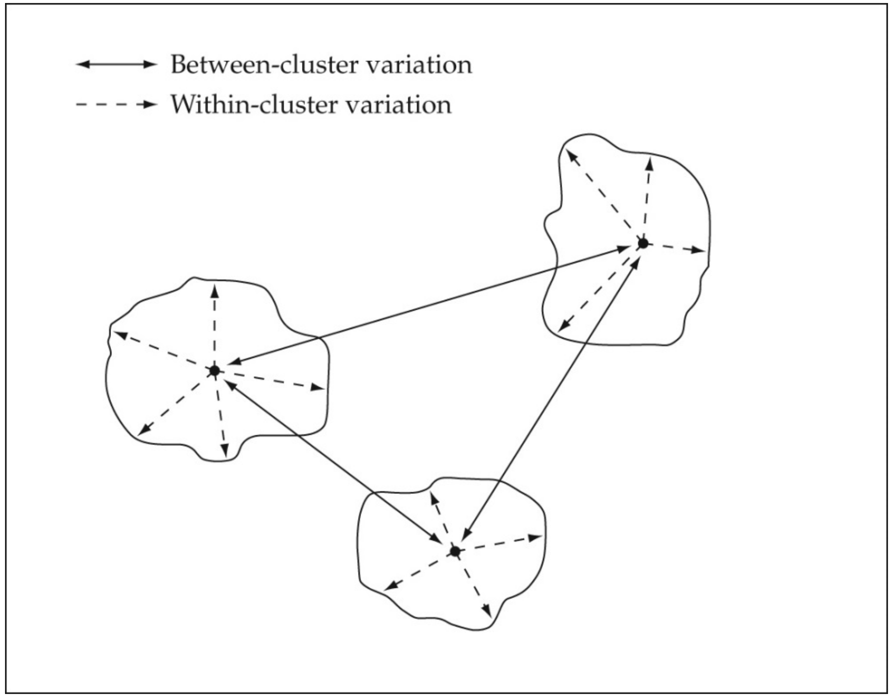
2. What kind of strategy do we follow in the amalgamation procedure (formation of clusters)?

3. How many clusters do we use

Yet,...1 universal aim

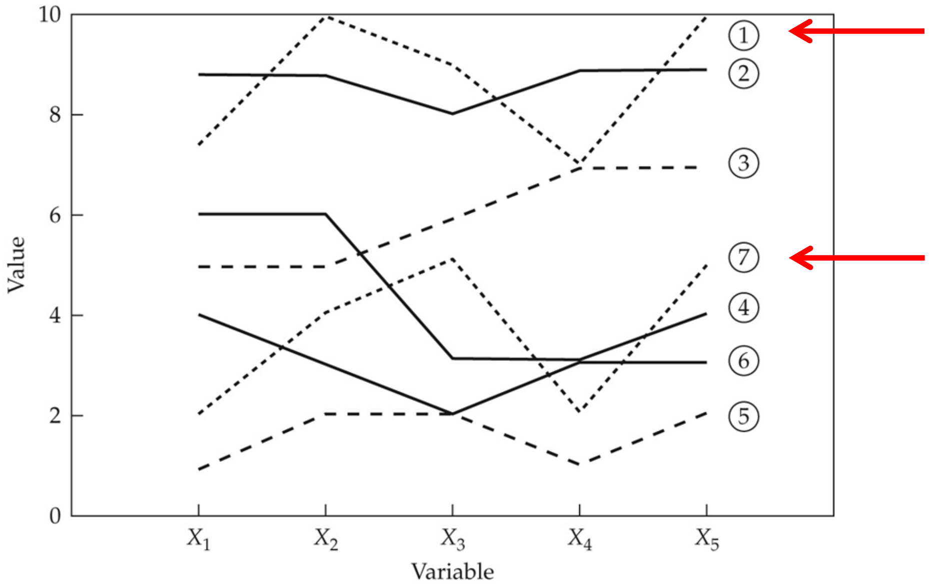
- Obtaining a limited number of mutually exclusive clusters

- Maximizing internal homogeneity (within cluster variation) and maximal external heterogeneity (between cluster variation)



Question 1a: Which similarity measure?

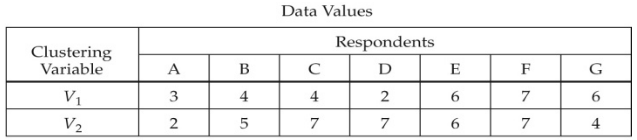
- Possibility A: measures of association or correlation
  - Focus on similarity of pattern of the scores, NOT on the level of the scores
  - Suitable for nominal/ordinal measurement level



Question 1b: Which similarity measure?

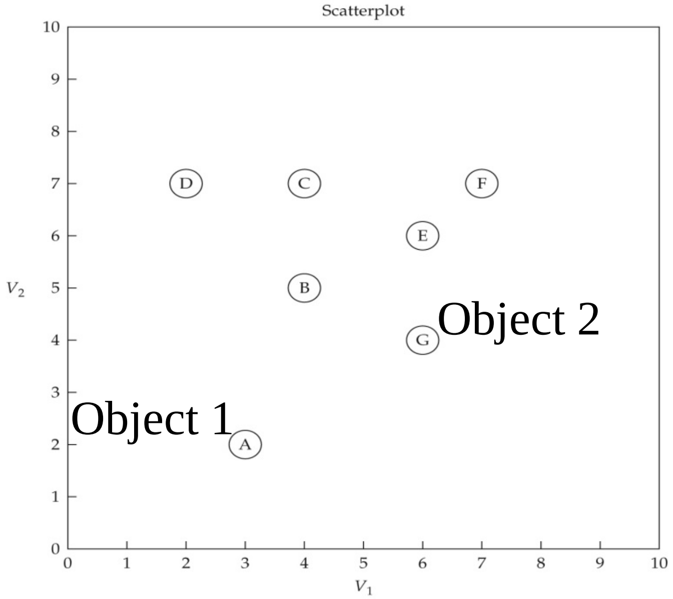
- Possibility A: distance measures (proximities, often based on Euclidian distance)
  - Focus on level of the scores, NOT on the pattern of the scores
  - Suitable for quantitative measurement level

- Squared Euclidean distance
- City-block (Manhattan) distance
- Chebychev distance
- Mahalanobis distance (D2)
(for Multicollinear variables)



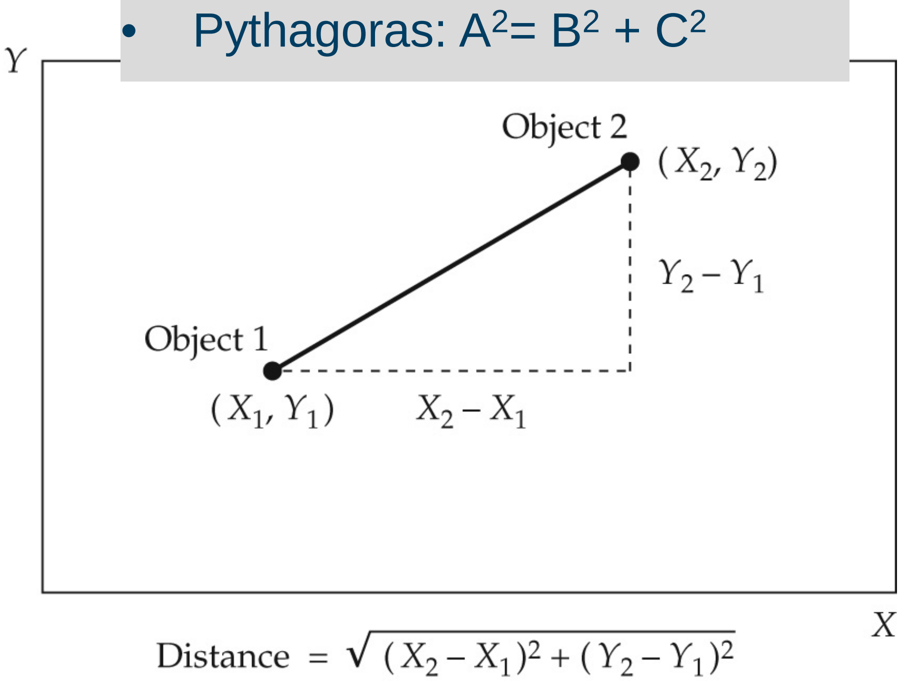
Data Values

Clustering Variable	Respondents						
	A	B	C	D	E	F	G
V_1	3	4	4	2	6	7	6
V_2	2	5	7	7	6	7	4



Scatterplot

- Pythagoras: $A^2 = B^2 + C^2$



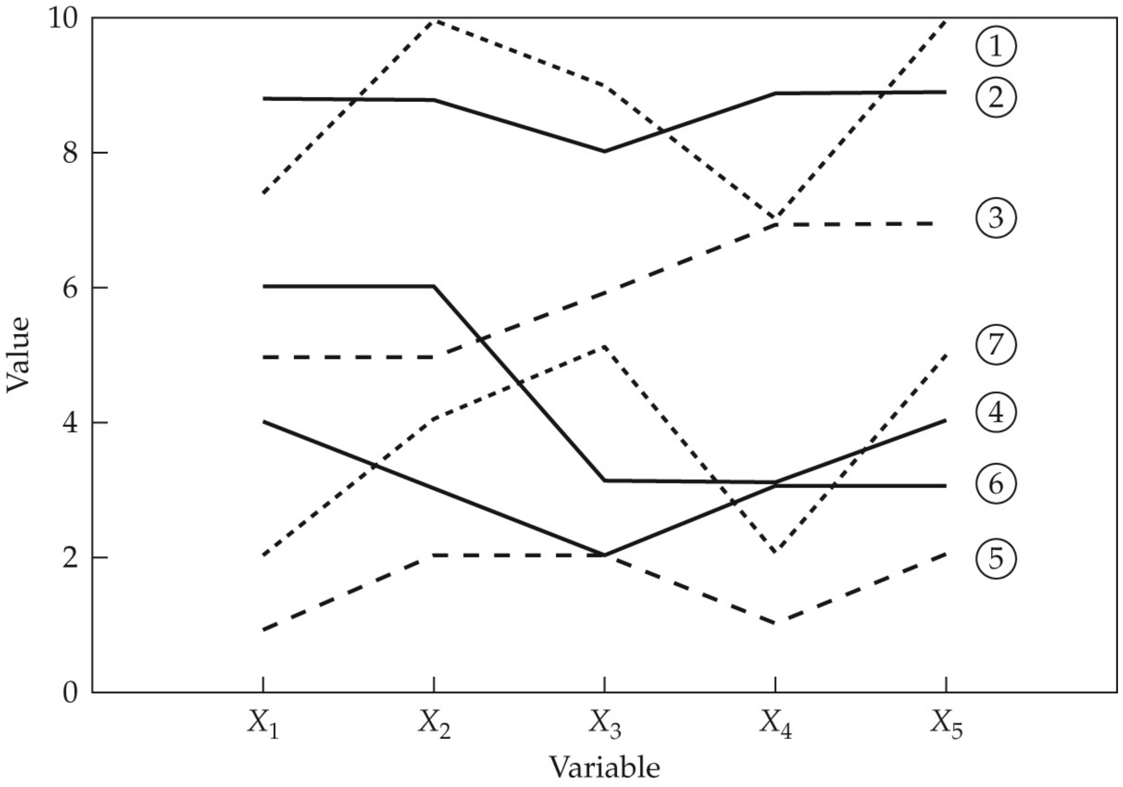
$$\text{Distance} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

Dissimilarity matrix in SPSS

	Squared Euclidean Distance				
Case	1	2	3	4	5
1		325,000	425,000	500,000	50,000
2	325,000		200,000	125,000	125,000
3	425,000	200,000		25,000	225,000
4	500,000	125,000	25,000		250,000
5	50,000	125,000	225,000	250,000	

Question 1c: Which similarity measure?

- Correlation versus distance



Smallest distance:
- between 1 and 2

Highest distance:
- between 1 and 5
- between 2 and 5

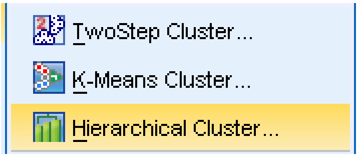
Highest correlation:
- between 1 and 5
- between 1 and 7

Never forget: use standardised variables to compute distance

$$z_i = \frac{(x_i - \bar{x})}{s}$$

Differences in measurement unit have a strong impact on the formation of cluster

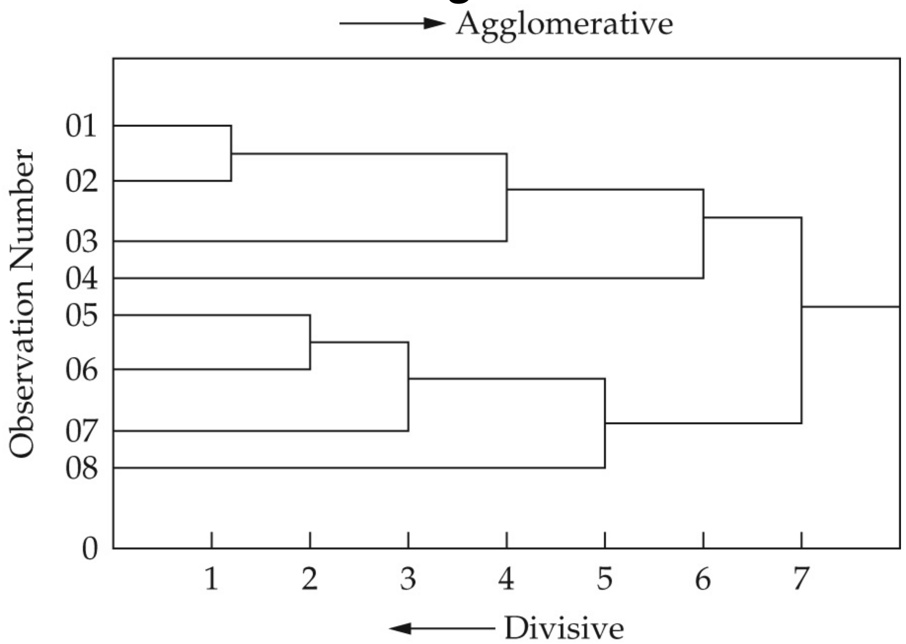
Question 2: Which agglomeration schedule



TwoStep Cluster...
K-Means Cluster...
Hierarchical Cluster...

- Hierarchical cluster methods
  - Clusters are nested
  - agglomerative (bottom-up) versus divisive (top-down)
  - Time – and labour intensive

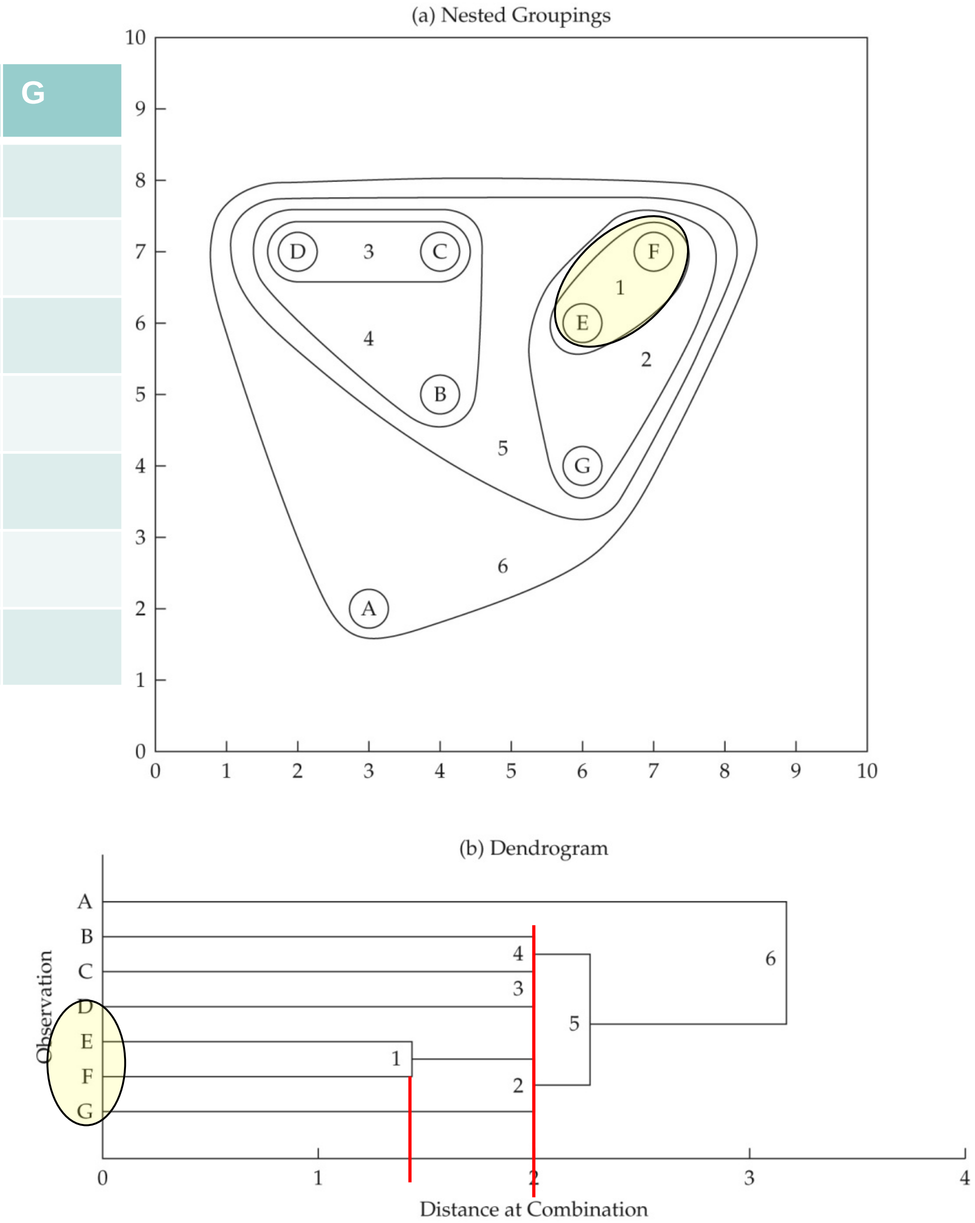
- Non-hierarchical cluster methods
  - Clusters are non-nested e.g. SPSS ‘*K-means clustering*’
  - Less costly
  - Iterative process based on. ‘*seeds*’



Nested structure of hierarchical clustering

	A	B	C	D	E	F	G
A							
B	3,162						
C	5,099	2,000					
D	5,099	2,828	2,000				
E	5,000	2,236	2,236	4,123			
F	6,403	3,606	3,000	5,000	1,414		
G	3,606	2,236	3,606	5,000	2,000	3,162	

Internal homogeneity of the clusters decreases in each consecutive step (mean distance in clusters)


(a) Nested Groupings

(b) Dendrogram

Question 2: Hierarchical agglomeration methods

- Nearest (single linkage) versus Farthest neighbour (complete linkage) procedure
- Between-groups linkage (average linkage)
- Within-groups linkage

Best buy

- Ward's method
  (min. Sum of Squares (SS) of each cluster pair that can be formed in each step)
- Centroïd method

Use Squared-Euclidean distance

Question 3: optimal number of clusters

- Grafical: dendogram or icicle-plot
- Numerical: ‘agglomeration schedule’
  (based on a strong increase in within-cluster distance)
- Theoretical: external and predictive validation

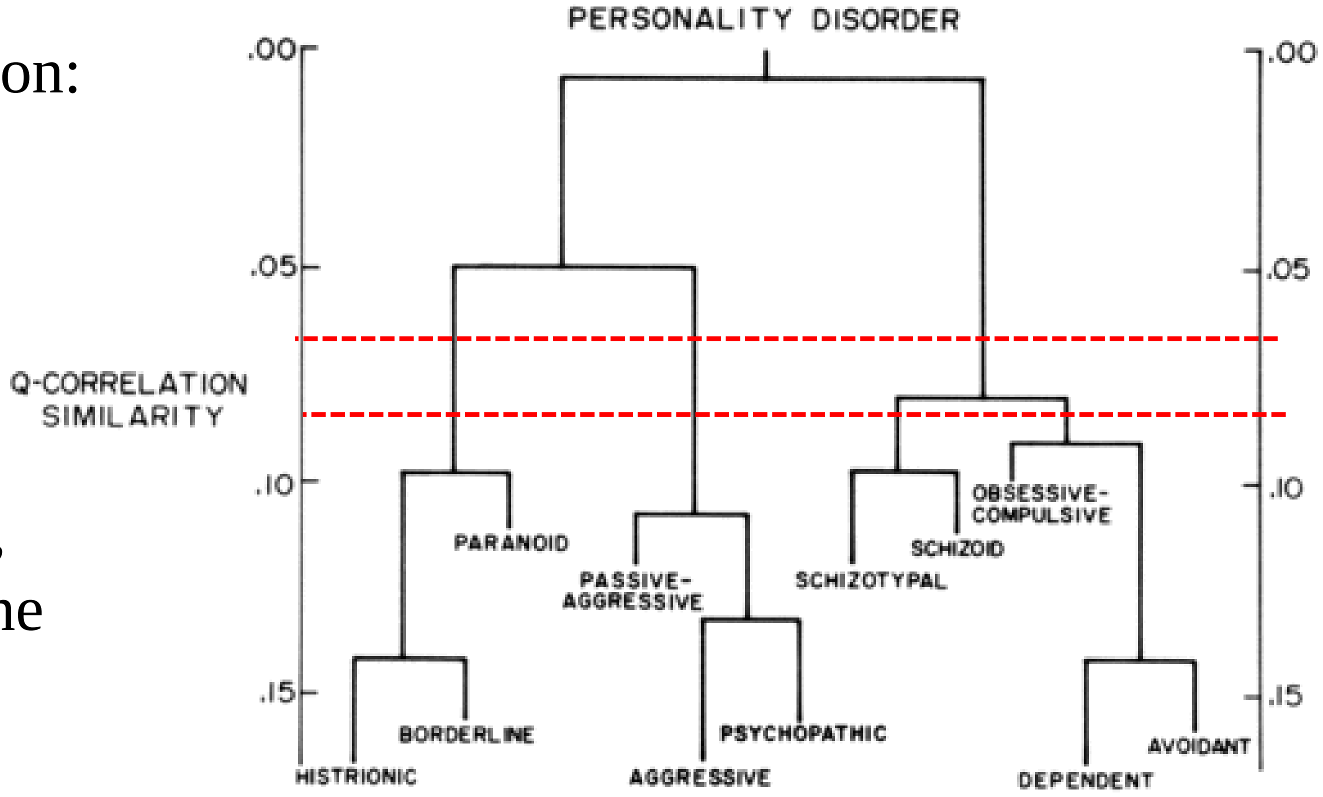
Vraag 3: het optimale # clusters

- Grafical: dendogram or icicle-plot

Standard classification:
Cluster A:
Paranoid, schizoid, schizotypal

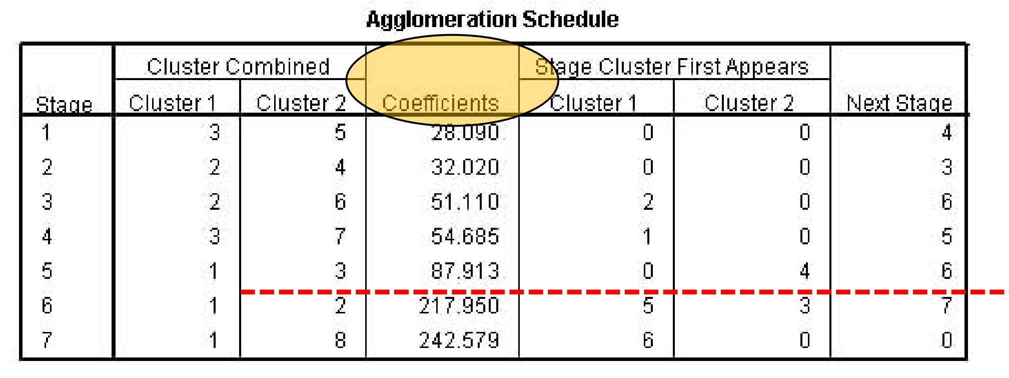
Cluster B:
Theatral, narcistical, anti-social, borderline

Cluster C:
Avoidant, dependent, obsessive-compulsive



Question 3: optimal number of clusters

- Numerical: ‘agglomeration schedule’
  (strong increase qua within-cluster distance)

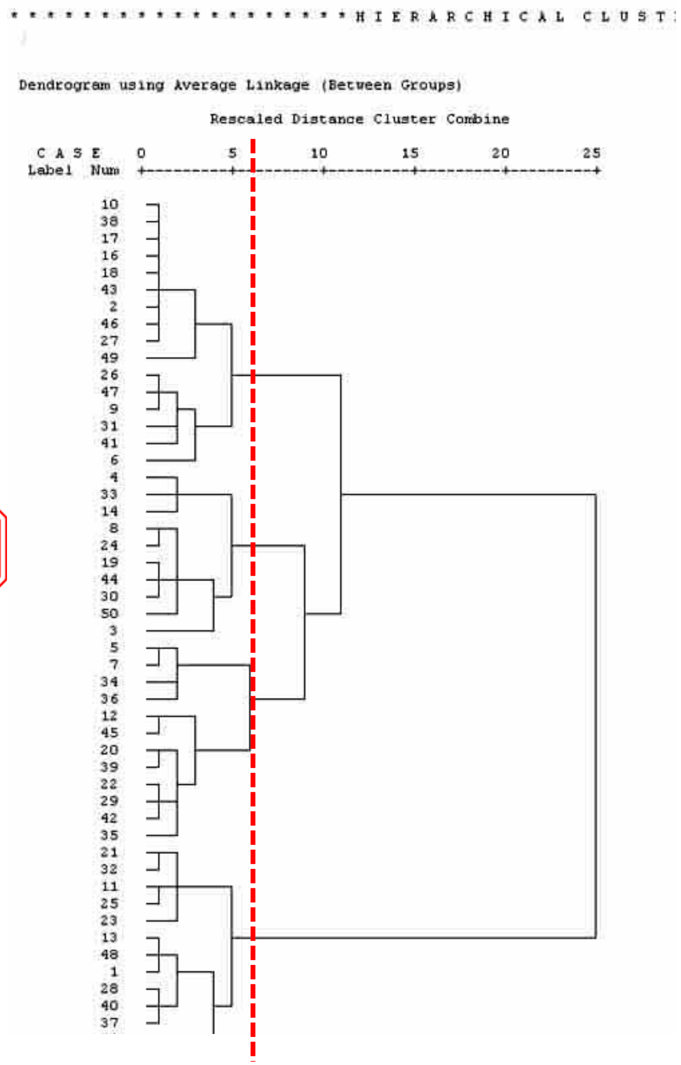


Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	5	28.090	0	0	4
2	2	4	32.020	0	0	3
3	2	6	51.110	2	0	6
4	3	7	54.685	1	0	5
5	1	3	87.913	0	4	6
6	1	2	217.950	5	3	7
7	1	8	242.579	6	0	0

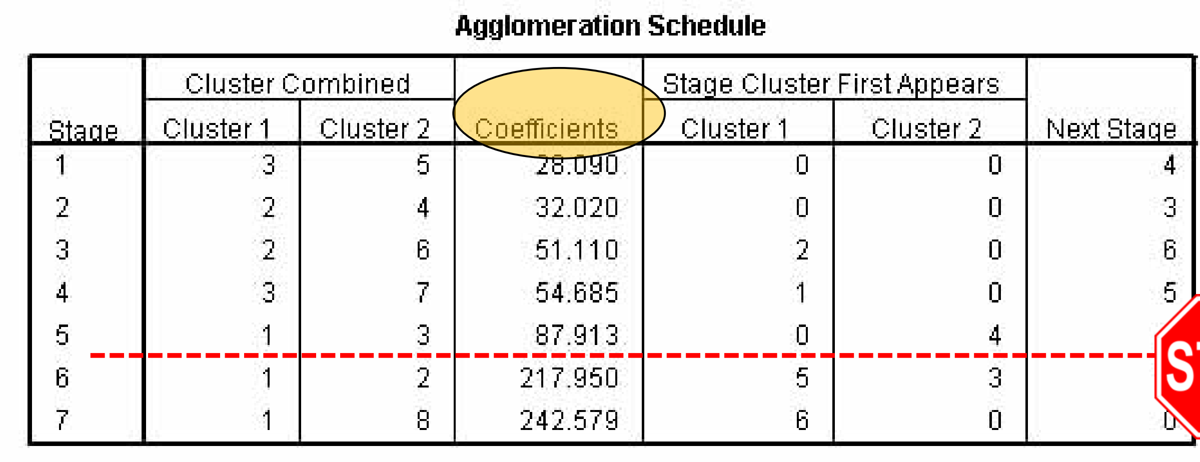
big jump=
strongly dissimilar
clusters are
agglomerated

STOP



Question 3: optimal number of clusters

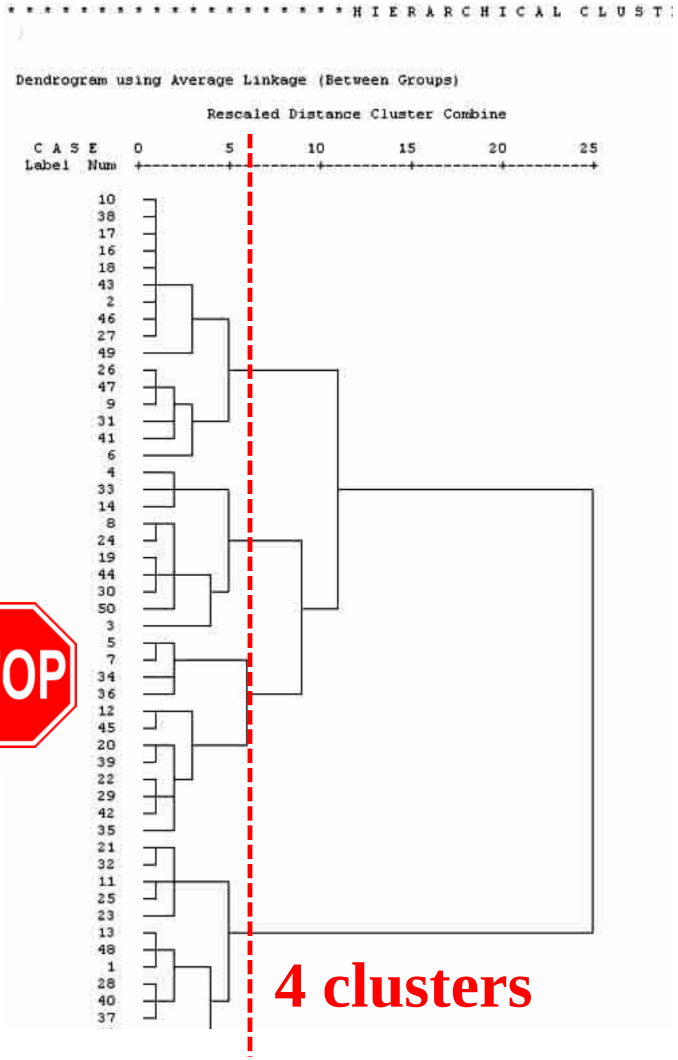
Comparison: dendrogram - agglomeration



Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	5	28.090	0	0	4
2	2	4	32.020	0	0	3
3	2	6	51.110	2	0	6
4	3	7	54.685	1	0	5
5	1	3	87.913	0	4	6
6	1	2	217.950	5	3	7
7	1	8	242.579	6	0	0

STOP


* * * * * * * * * * * * * * * * * * H I E R A R C H I C A L C L U S T
Dendrogram using Average Linkage (Between Groups)
Rescaled Distance Cluster Combine
C A S E
Label Num
0
5
10
15
20
25
10
38
17
16
18
43
2
46
27
49
26
47
9
31
41
6
4
33
14
8
24
19
44
30
50
3
5
7
34
36
12
45
20
39
22
29
42
35
21
32
11
25
23
13
48
1
28
40
37

4 clusters