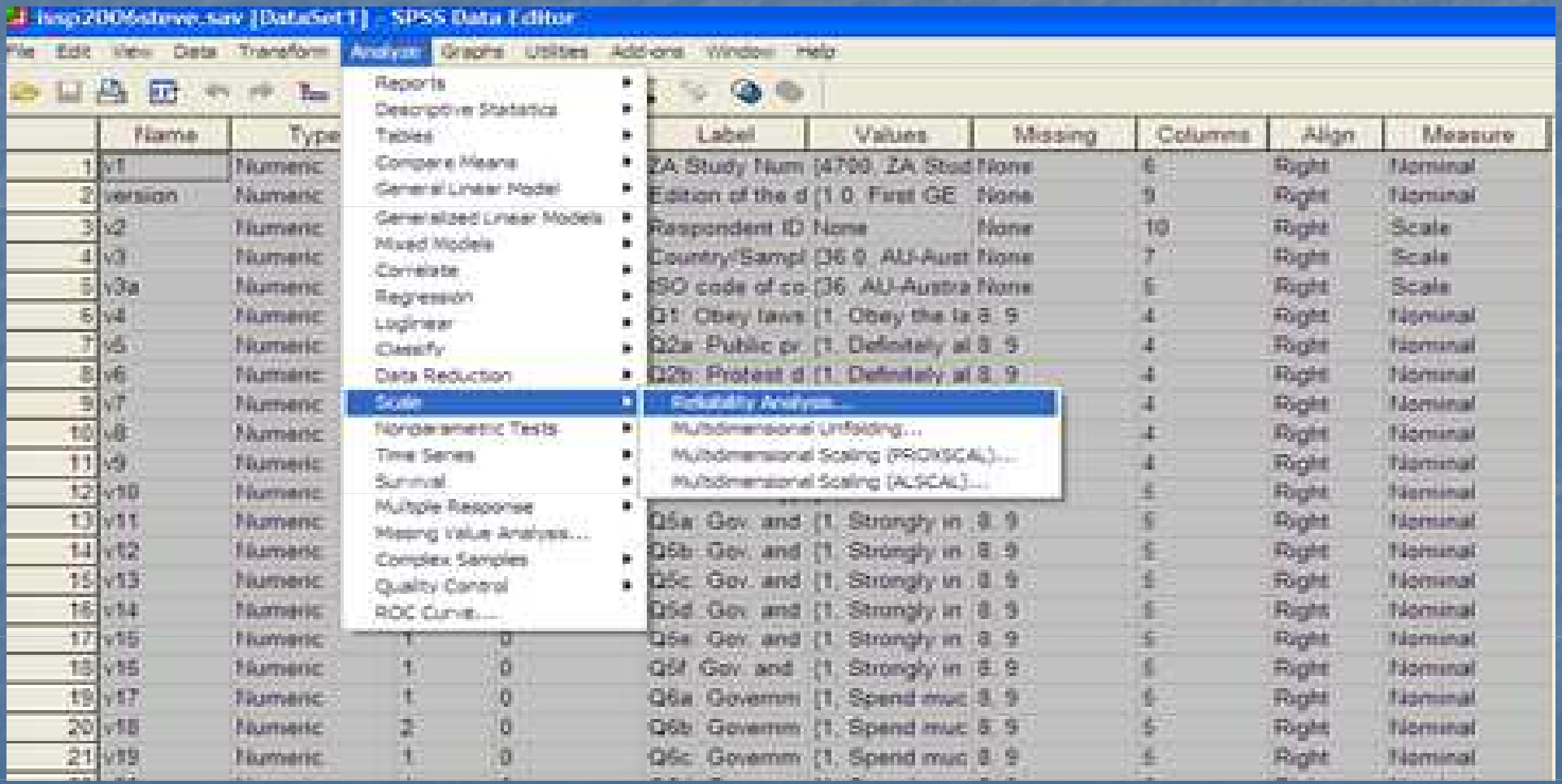
Scaling Lab

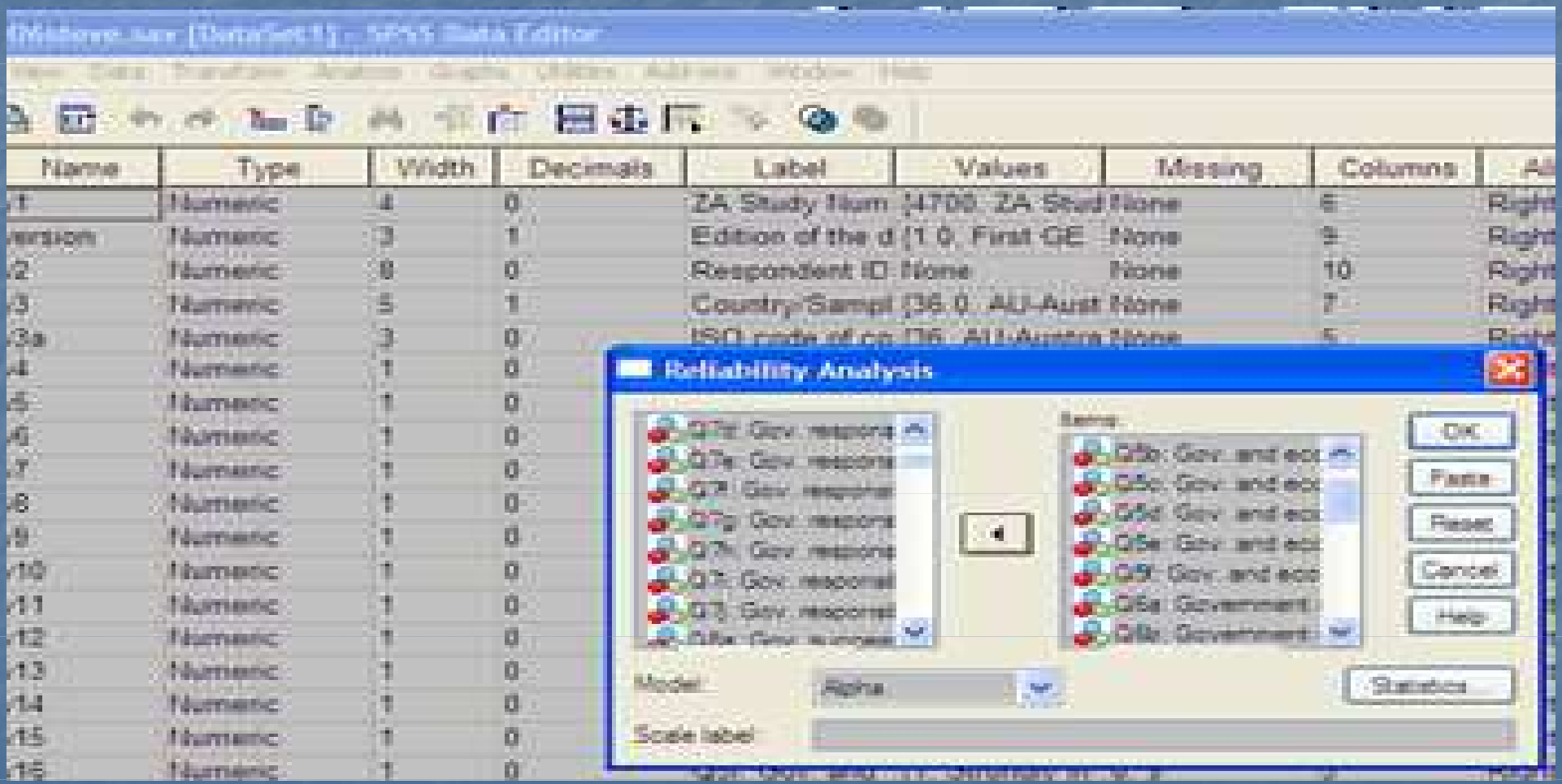
Scaling outline

- First add the country filter
- Then we will do Cronbach's Alpha
- Then Factor Analysis

Starting to make a one-dimensional scale with Cronbach's Alpha



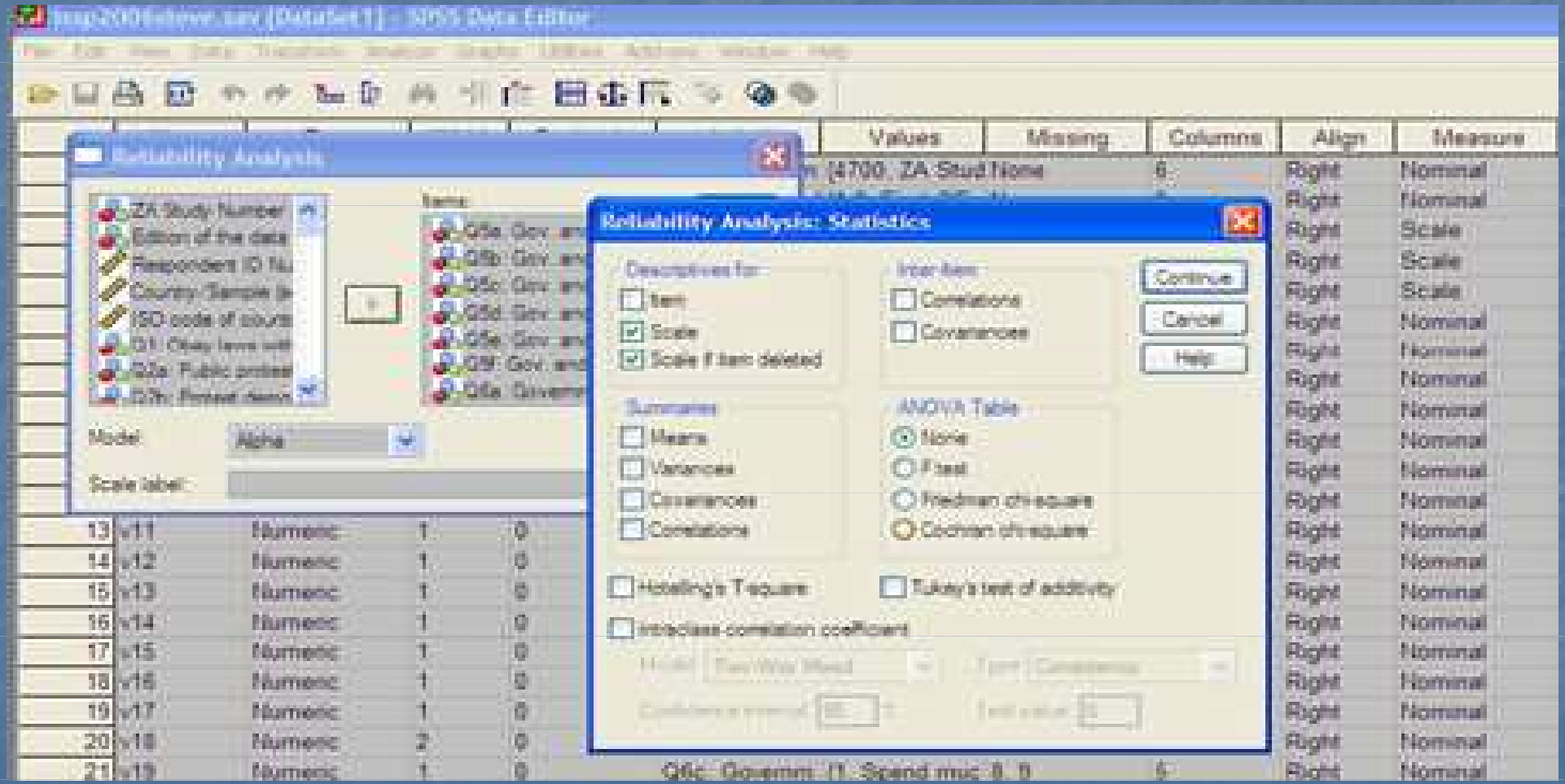
Add the items



In the present example the score will be low

- Why?
- I did not RECODE the variables
- They should all go in the same direction so that the most points is either for government intervention or against it

Click on ”Statistics.”Choose ”scale” and ”scale if item deleted”



Alpha Score in SPSS. Not so bad since in this example I did not recode the variables, so they all go in the same direction

Reliability Statistics

Cronbach's Alpha	N of Items
.675	16

We should delete variables that lower Alpha, which means Alpha>.675 if Alpha item Deleted. We see that the first question Q5a would increase Alpha if eliminated and that its correlation is negative (perhaps because it was not recoded?)

Item-Total Statistics: Scale ---- if item Deleted

	Mean iID	Var iID	CorrItTotCor	Alpha iIDeleted
Q5a: Gov. and economy: Cuts in gov. spending	38.48	38.062	-.037	.704
Q5b: Gov. and economy: Financing projects for new jobs	38.79	33.260	.417	.643
Q5c: Gov. and economy: Less gov. reg. of business	38.51	36.464	.103	.684
Q5d: Gae: Support industry to develop new products	38.58	33.468	.376	.648
Q5e: Gae: Support declining industries to protect jobs	37.87	31.666	.431	.637
Q5f: Gov. and economy: Red. working week for more jobs	37.71	34.531	.226	.669
Q6a: Government should spend money: Environment	38.33	34.922	.328	.656
Q6b: Government should spend money: Health	38.76	34.806	.345	.654
Q6c: Government should spend money: Law enforcement	38.04	34.856	.305	.658
Q6d: Government should spend money: Education	38.66	35.122	.305	.658
Q6e: Government should spend money: Defence	37.39	35.746	.210	.669
Q6f: Government should spend money: Retirement	38.56	34.101	.389	.648
Q6g: Government should spend money: Unempl. benefits	37.56	34.475	.306	.657
Q6h: Government should spend money: Culture and arts	37.78	35.236	.300	.659
Q7a: Gov. responsibility: Provide job for everyone	38.94	34.990	.268	.662
Q7b: Gov. responsibility: Control prices	38.60	35.019	.265	.662

After eliminating we get a better score

Reliability Statistics

Cronbach's Alpha	N of Items
.704	15

Now Alpha could be increased if we take away Q5c

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q5b: Gov. and economy: Financing projects for new jobs	36.42	32.748	.402	.678
Q5c: Gov. and economy: Less gov. reg. of business	36.13	35.952	.086	.717
Q5d: Gov. and economy: Support industry to develop new products	36.20	32.818	.373	.681
Q5e: Gov. and economy: Support declining industries to protect jobs	35.49	30.707	.456	.668
Q5f: Gov. and economy: Red. working week for more jobs	35.32	33.695	.237	.700
Q6a: Government should spend money: Environment	35.94	34.253	.325	.688
Q6b: Government should spend money: Health	36.37	34.074	.348	.685
Q6c: Government should spend money: Law enforcement	35.66	34.262	.294	.691
Q6d: Government should spend money: Education	36.28	34.317	.317	.689
Q6e: Government should spend money: Defence	35.00	35.095	.206	.701
Q6f: Government should spend money: Retirement	36.18	33.351	.395	.680
Q6g: Government should spend money: Unempl. benefits	35.17	33.541	.328	.687
Q6h: Government should spend money: Culture and arts	35.41	34.396	.315	.689
Q7a: Gov. responsibility: Provide job for everyone	36.55	34.037	.293	.691
Q7b: Gov. responsibility: Control prices	36.22	34.109	.285	.692

Alpha increases once more!

Reliability Statistics

Cronbach's Alpha	N of Items
.716	14

Now Alpha cannot be increased by removing an item, as it would be less than .716 if any were removed

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q5b: Gov. and economy: Financing projects for new jobs	34.05	31.021	.367	.696
Q5d: Gov. and economy: Support industry to develop new products	33.82	31.303	.320	.702
Q5e: Gov. and economy: Support declining industries to protect jobs	33.13	28.525	.471	.680
Q5f: Gov. and economy: Red. working week for more jobs	32.96	31.615	.234	.715
Q6a: Government should spend money: Environment	33.57	32.159	.318	.702
Q6b: Government should spend money: Health	34.00	31.877	.358	.698
Q6c: Government should spend money: Law enforcement	33.29	31.993	.307	.703
Q6d: Government should spend money: Education	33.90	32.337	.301	.704
Q6e: Government should spend money: Defence	32.64	32.919	.207	.714
Q6f: Government should spend money: Retirement	33.82	31.076	.415	.691
Q6g: Government should spend money: Unempl. benefits	32.81	31.101	.358	.697
Q6h: Government should spend money: Culture and arts	33.03	32.263	.314	.702
Q7a: Gov. responsibility: Provide job for everyone	34.19	31.638	.321	.701
Q7b: Gov. responsibility:				

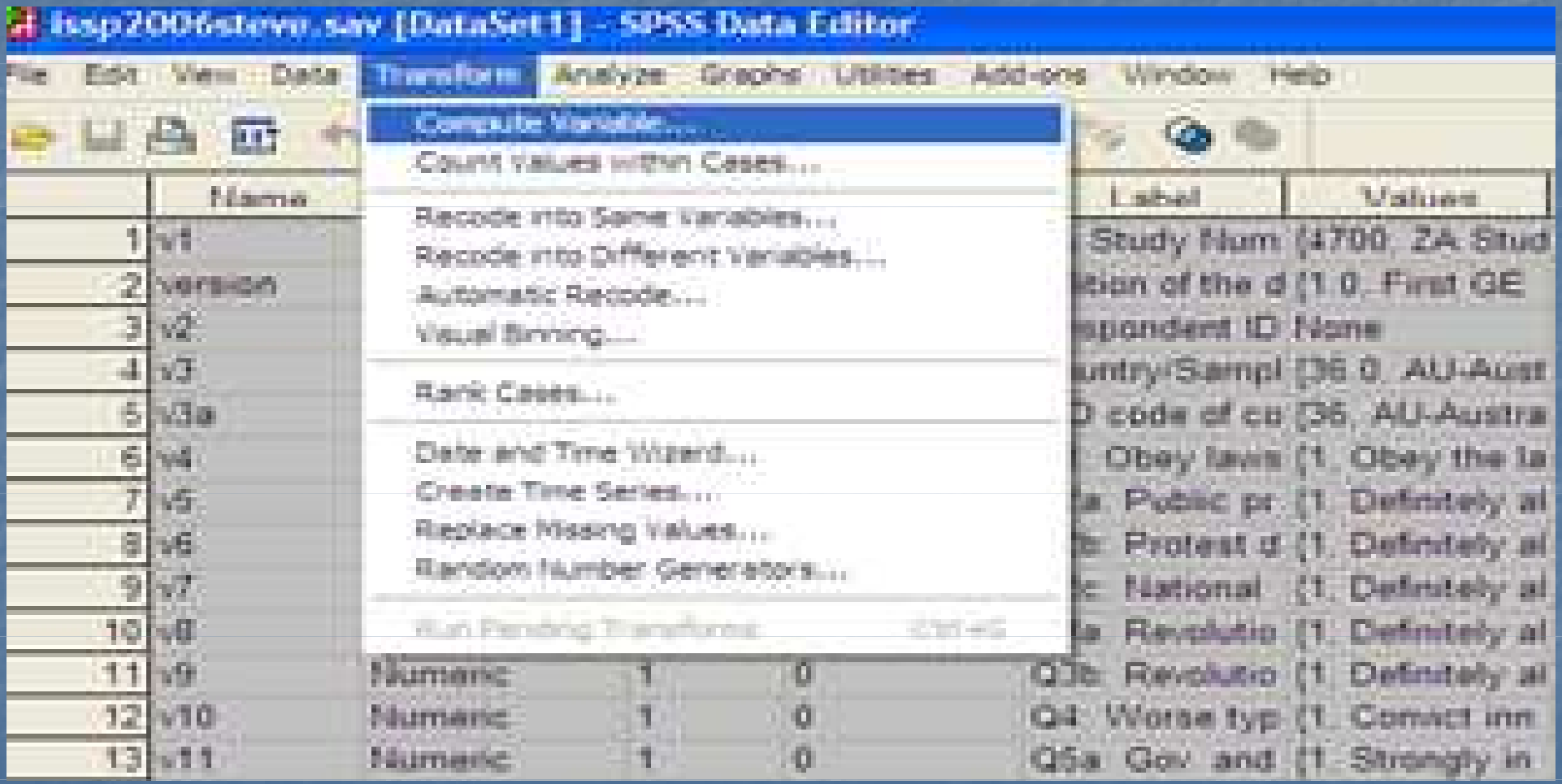
One little problem....

- If we are comparing several groups, like men and women or Sweden and France, then we could compare the means and say that French score higher than Swedes or Women score higher than men
- Because I was lazy in making this presentation, I did not recode the questions 7a and 7b, but since they are on a scale of 1-4, while the others are on a scale of 1-5, I really should have transformed them by multiply the scores by 1.25, so they too would have the same scale.
- There would still be a problem, because even though they would all have the same maximum value, they would not have the same minimum.
- The best would be to rescale ALL the variables used, so that instead of 1-5 they would be 0, 2,3,4
- Then rescale the 1-4 scale so the scale would be 0, 1.33, 2.67, 4.0

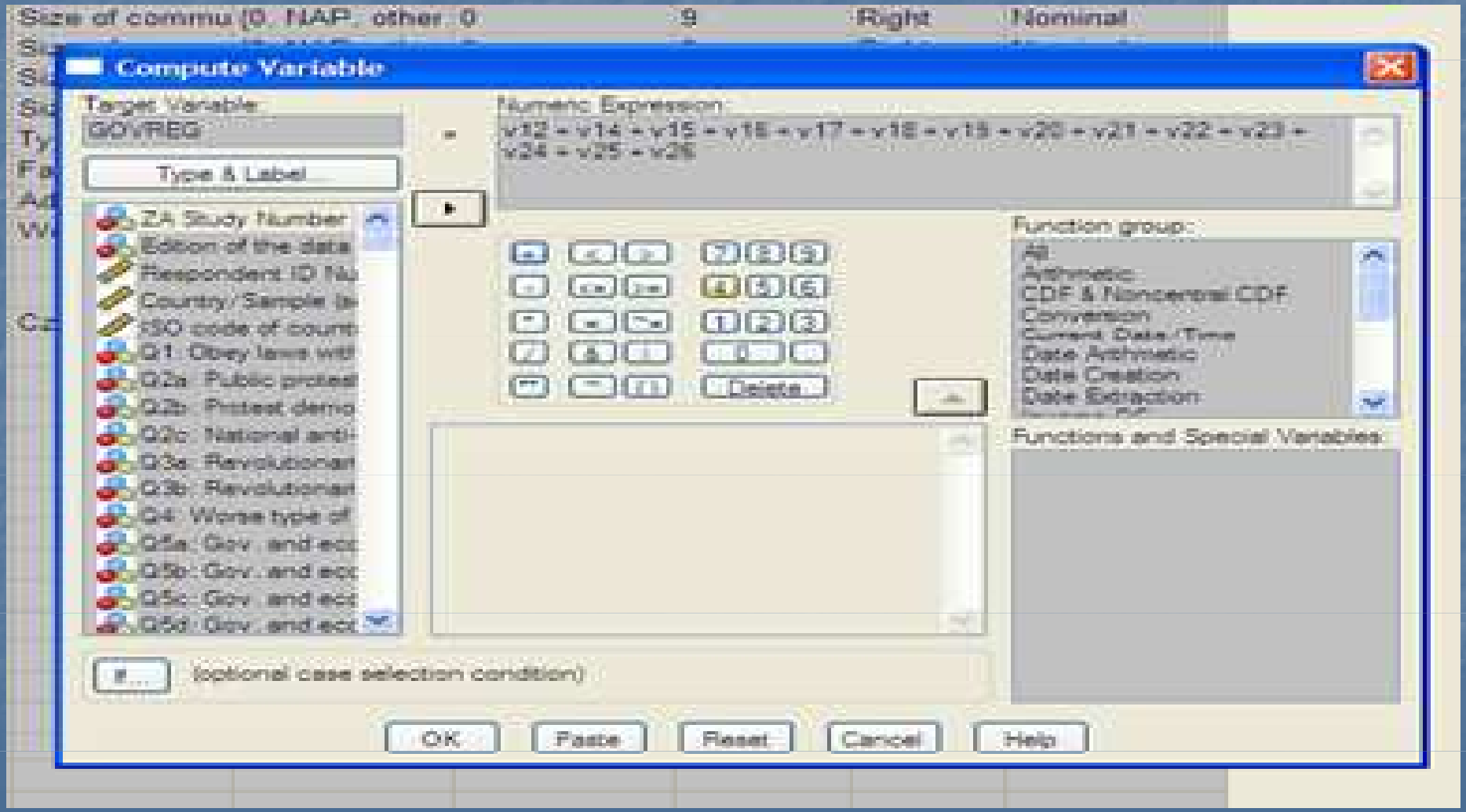
The scale scores

- Once we made these transformations so that all questions have a scale of 0-4
- AND they all go in the same direction, so that 4 for EVERY question either means support for government intervention or opposition to government intervention
- THEN we can compare the average scores among groups
- There are 14 questions in our scale, if all questions are from 0-4, then the scale would be from 0-56

Making the Scale in SPSS: Go back to compute variable



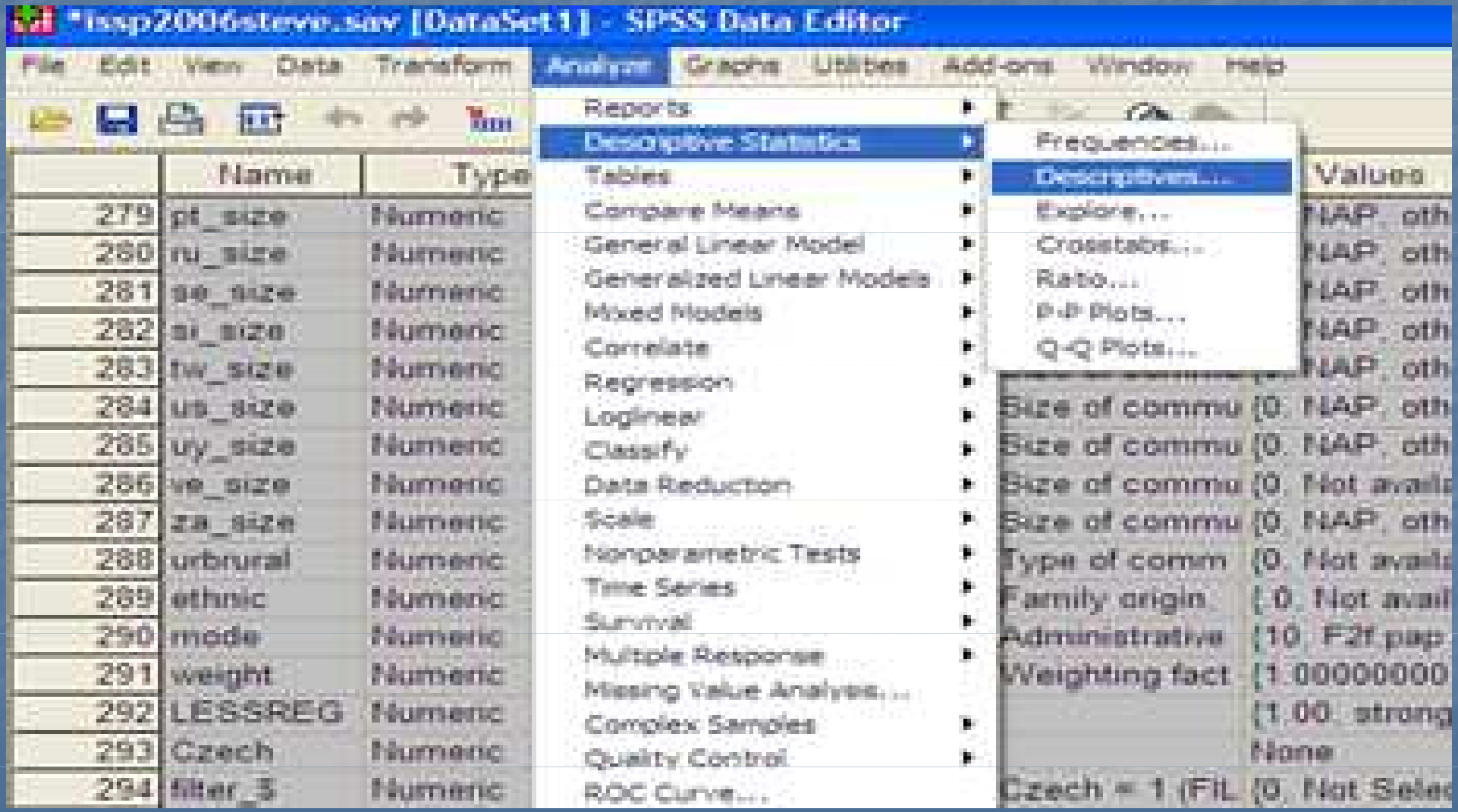
Make a new variable by adding the items together than comprise the Alpha score (that is minus Q5a and Q5c)



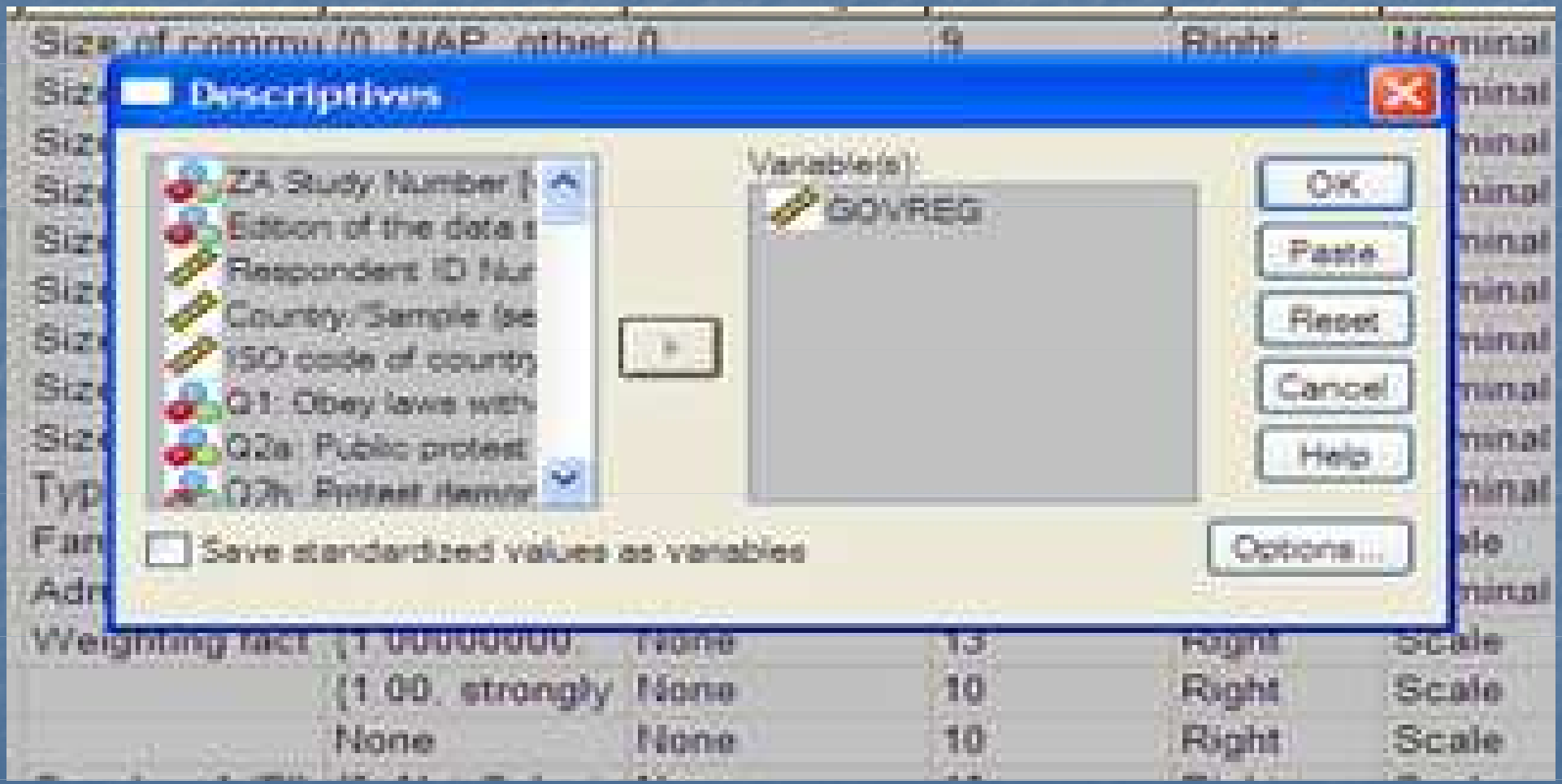
Your alpha assignment

- Choose the variables that you will want to include in your scale.
- Make sure that you have recoded variables so that each question has the same scale (like 0-4)
- Make sure that each question is also scaled in the same direction. You should have already done this during previous computer labs.
- Run the alpha analysis
- Eliminate variables if you can improve the score
- Create a scale
- Compare the means for two groups, like for Czechs and Swedes or for men and women. You can do this by making a filter for each group (like Czech and Swedes).
- Then apply the filter for one group (Czech) and go to ANALYZE->DESCRIPTIVE STATISTICS -> DESCRIPTIVES

Like this....


*lssp2006steve.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Quality Control
ROC Curve...
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...
Name
Type
Values
279 pt_size Numeric
280 ru_size Numeric
281 se_size Numeric
282 si_size Numeric
283 tw_size Numeric
284 us_size Numeric
285 uy_size Numeric
286 ve_size Numeric
287 za_size Numeric
288 urbrural Numeric
289 ethnic Numeric
290 mode Numeric
291 weight Numeric
292 LESSREG Numeric
293 Czech Numeric
294 filter_$ Numeric

Click on OK


Descriptives
Variable(s):
GOVREG
OK
Paste
Reset
Cancel
Help
Options...
Save standardized values as variables

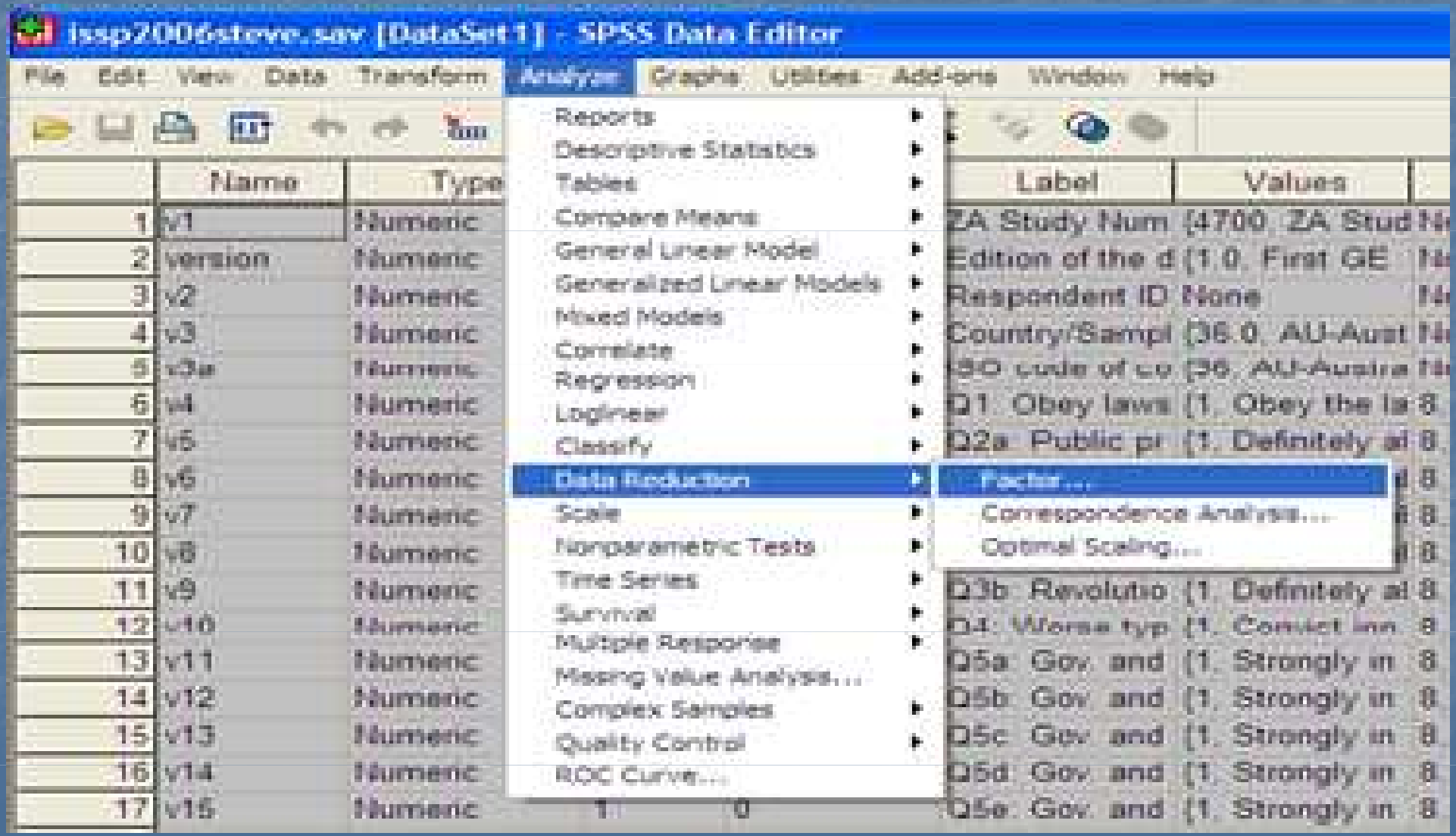
You can see the average score now (the mean)

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
GOVREG	993	19.00	63.00	36.0806	5.98280
Valid N (listwise)	993				

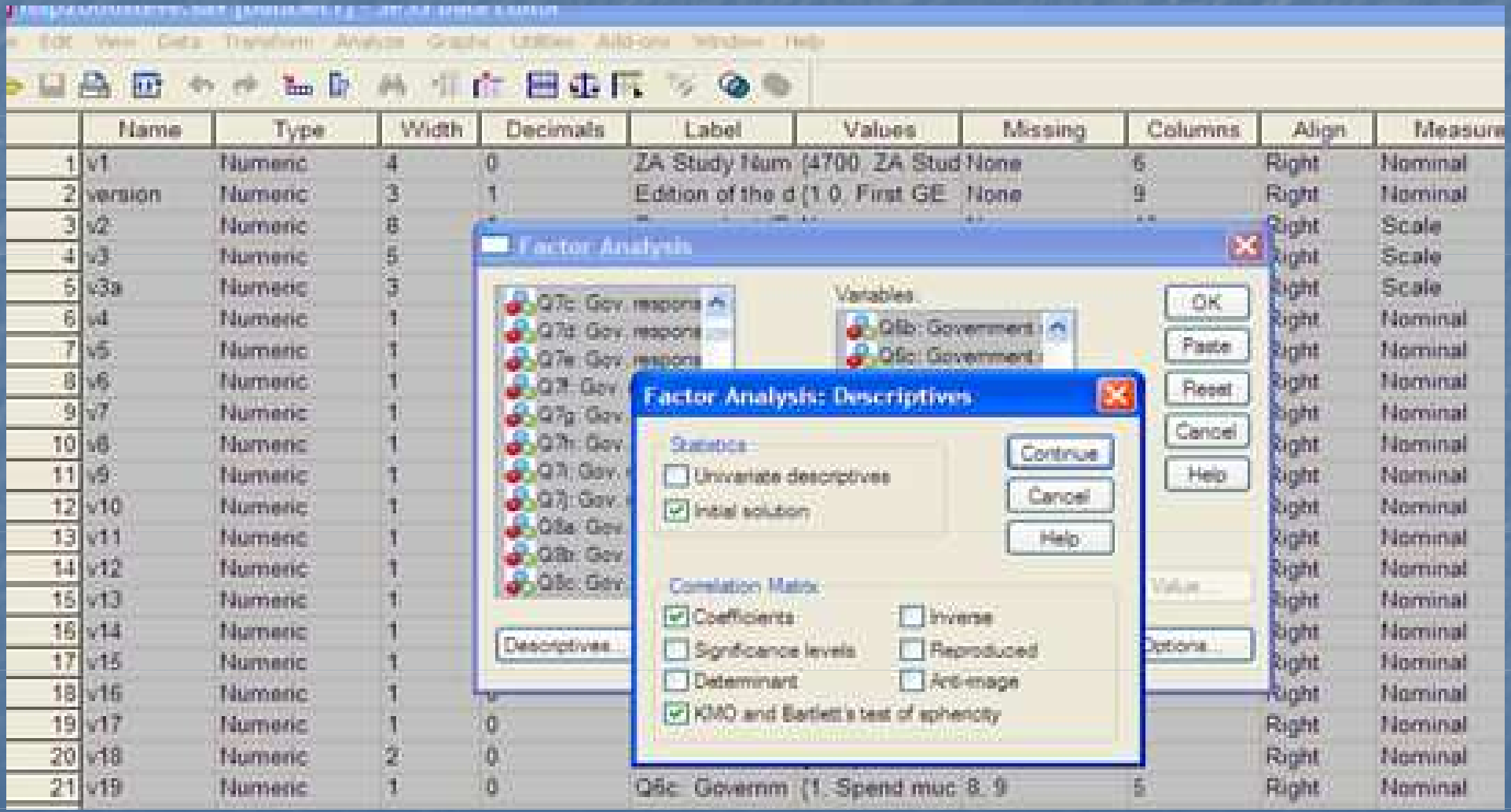
After checking the alpha scale you can run multiple regressions on it

Factor Analysis

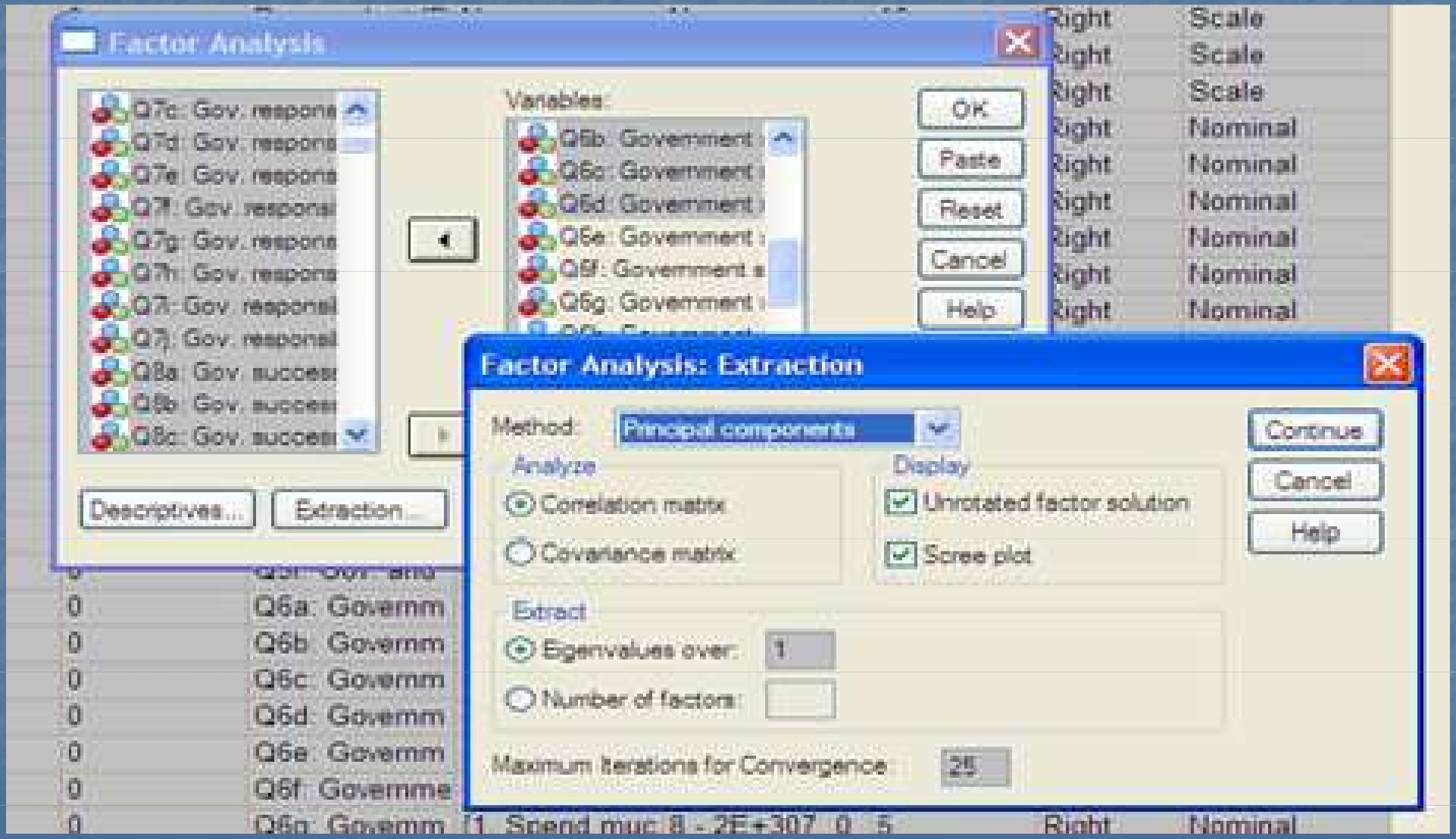


Choose the same variables that you chose for the first Cronbach Alpha calculation

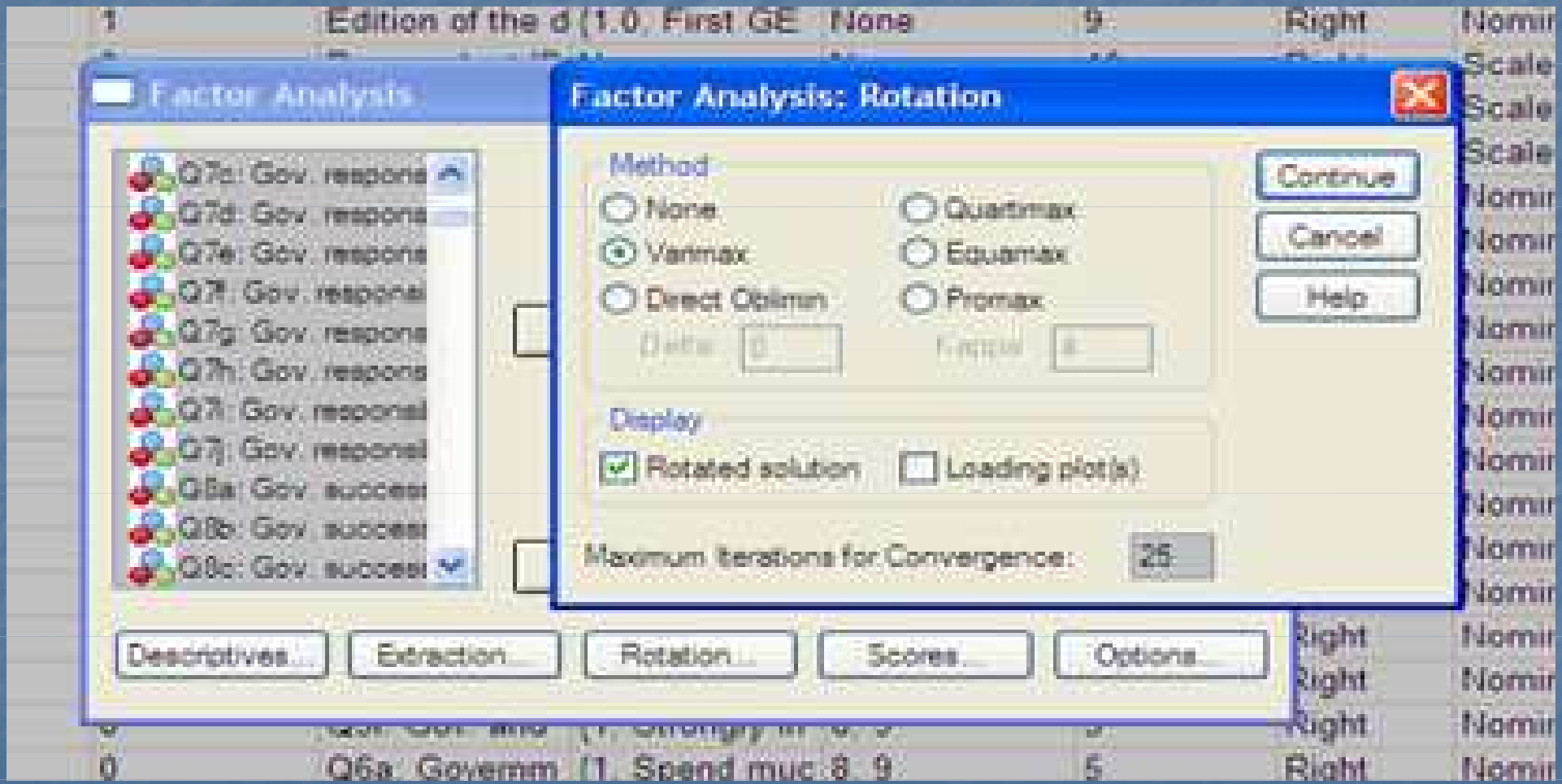
Choose the following in DESCRIPTIVES



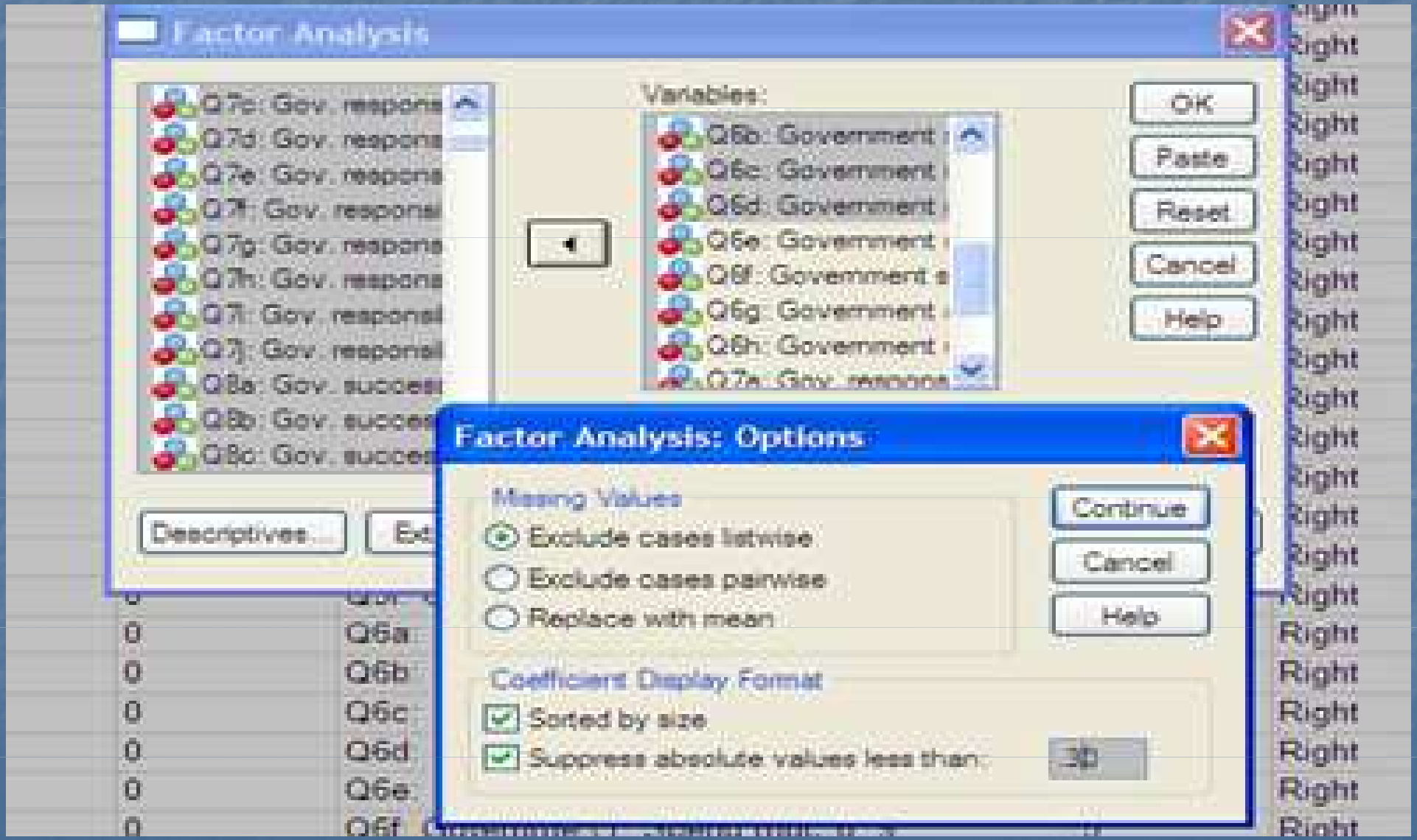
Go to EXTRACTION and click on SCREE PLOT



Then for ROTATION choose VARIMAX

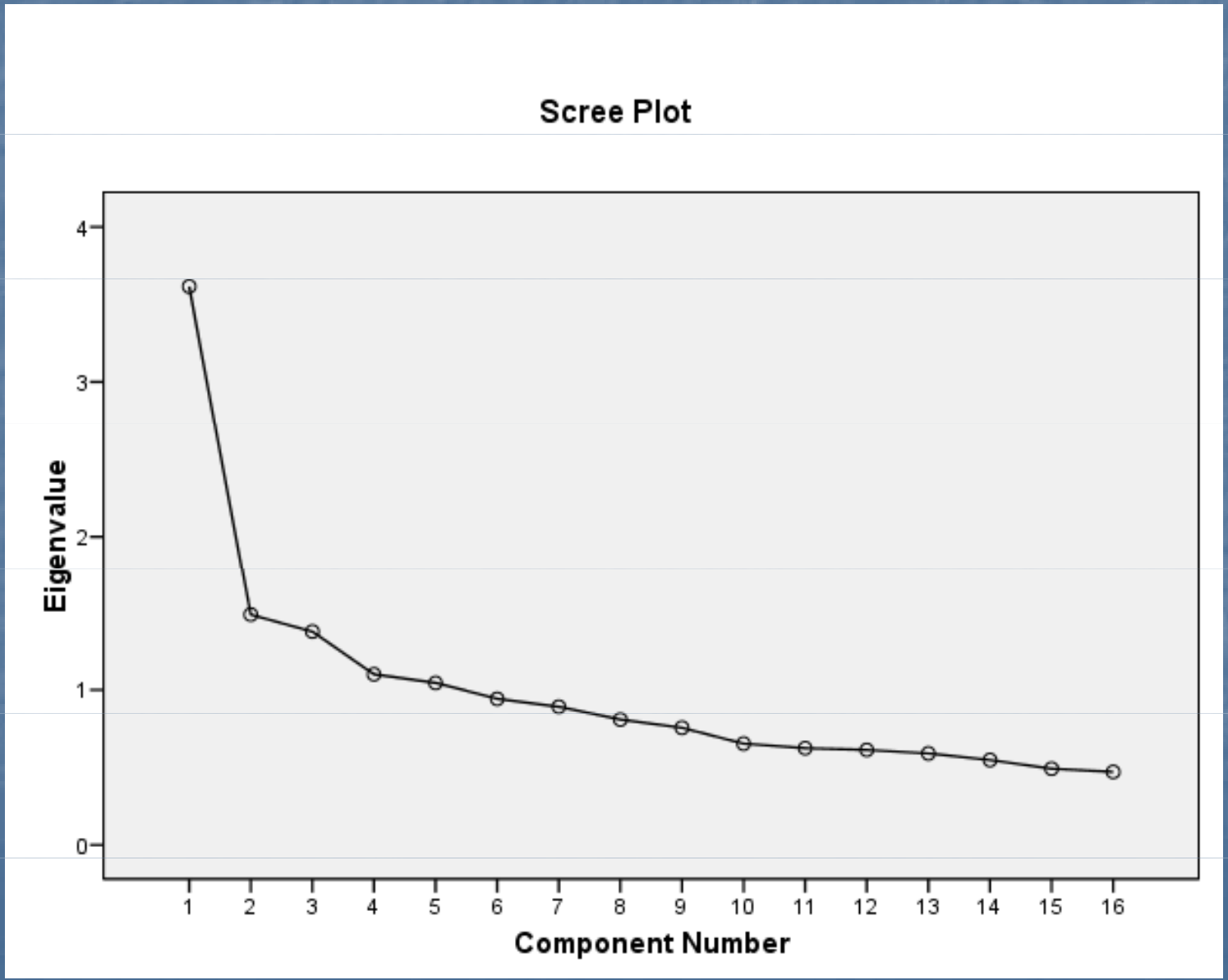


Go to OPTIONS and choose SORTED BY SIZE and under SUPRESS ABOLUTE VALUES change it to .30



Click on OK and you get your first result

First let's look at the scree plot



Analyzing the Scree plot

- 4 have values above 1
- But there seems to be a big dip after two components, which indicates that perhaps there are two dimensions

Let's Look at the explained variance: the first two components explain the most, although the next three do have Eigen values slightly more than 1. However, the third component does have a relatively high eigenvalue. Whether or not to keep the third factor depends partially on whether it makes theoretical sense.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.612	22.574	22.574	3.612	22.574	22.574	2.289	14.307	14.307
2	1.483	9.267	31.842	1.483	9.267	31.842	1.965	12.283	26.590
3	1.374	8.587	40.429	1.374	8.587	40.429	1.581	9.879	36.470
4	1.100	6.872	47.301	1.100	6.872	47.301	1.523	9.516	45.986
5	1.045	6.528	53.830	1.045	6.528	53.830	1.255	7.844	53.830
6	.943	5.892	59.721						
7	.892	5.574	65.296						
8	.808	5.052	70.348						
9	.756	4.722	75.070						
10	.653	4.083	79.153						
11	.624	3.898	83.050						
12	.613	3.829	86.879						
13	.590	3.686	90.565						
14	.547	3.422	93.987						
15	.492	3.072	97.059						
16	.471	2.941	100.000						

Extraction Method: Principal Component Analysis.

In the unrotated matrix many items are highly correlated to several factors

Component Matrix[a]

	Component				
	1	2	3	4	5
Q6f: Government should spend money: Retirement	.645				
Q6g: Government should spend money: Unempl. benefits	.631				.311
Q6b: Government should spend money: Health	.610				
Q6d: Government should spend money: Education	.581	-.335			
Q5e: Gov. and economy: Support declining industries to protect jobs	.545	.442			
Q7a: Gov. responsibility: Provide job for everyone	.536	.389			
Q6h: Government should spend money: Culture and arts	.531			-.364	.388
Q5b: Gov. and economy: Financing projects for new jobs	.530	.318			
Q7b: Gov. responsibility: Control prices	.475	.351			
Q6c: Government should spend money: Law enforcement	.410	-.389		.352	
Q5f: Gov. and economy: Red. working week for more jobs	.393	.303			.351
Q5c: Gov. and economy: Less gov. reg. of business			.662		.336
Q5d: Gov. and economy: Support industry to develop new products	.353		.524		-.359
Q5a: Gov. and economy: Cuts in gov. spending			.510		.410
Q6a: Government should spend money: Environment	.386	-.404		-.533	
Q6e: Government should spend money: Defence	.344	-.307		.465	

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

But in the rotated matrix there are only a few overlaps

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Q7a: Gov. responsibility: Provide job for everyone	.714				
Q6g: Government should spend money: Unempl. benefits	.608		.326		
Q7b: Gov. responsibility: Control prices	.589				
Q5e: Gov. and economy: Support declining industries to protect jobs	.562			.423	
Q5f: Gov. and economy: Red. working week for more jobs	.558				
Q6c: Government should spend money: Law enforcement		.728			
Q6e: Government should spend money: Defence		.637			
Q6b: Government should spend money: Health		.569			
Q6f: Government should spend money: Retirement	.453	.541			
Q6d: Government should spend money: Education		.500	.395		
Q6a: Government should spend money: Environment			.759		
Q6h: Government should spend money: Culture and arts			.744		
Q5d: Gov. and economy: Support industry to develop new products				.785	
Q5b: Gov. and economy: Financing projects for new jobs	.314			.675	
Q5c: Gov. and economy: Less gov. reg. of business					.758
Q5a: Gov. and economy: Cuts in gov. spending					.723

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

a. Rotation converged in 8 iterations.

Second Attempt

- I will remove the last 4 items, Q5d, Q5b, Q5c and Q5a, because they belong to components 4 and 5
- Please note: I am only making an example.
- If I were to do this seriously, I would first have to recode all the questions, so that they have the same scale (0-4 or 1-5) AND they must go in the same direction.
- It is very possible that these last 4 items really do fit in well, but they are coded in opposite directions – in some cases 5 denotes support for government intervention and in some cases it denotes opposition to intervention.

After removing these items, press OK

- Now there are only 3 factors.
- We see after rotation that some items are highly correlates with several components, so they should be eliminated.
- If Q6g and Q6e are eliminated, then we see that factor 3 makes sense: the culture and environment deal with quality of life rather than economic issues.
- Now we can eliminate as well Q6f because it is also highly correlated with two factors

Rotated Component Matrix[a]

	Component		
	1	2	3
Q7a: Gov. responsibility: Provide job for everyone	.724		
Q5e: Gov. and economy: Support declining industries to protect jobs	.653		
Q7b: Gov. responsibility: Control prices	.644		
Q6g: Government should spend money: Unempl. benefits	.546		.347
Q5f: Gov. and economy: Red. working week for more jobs	.540		
Q6c: Government should spend money: Law enforcement		.740	
Q6b: Government should spend money: Health		.599	
Q6e: Government should spend money: Defence		.598	
Q6d: Government should spend money: Education		.527	.423
Q6f: Government should spend money: Retirement	.431	.520	
Q6a: Government should spend money: Environment			.770
Q6h: Government should spend money: Culture and arts			.749

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

a. Rotation converged in 3 iterations.

New results: Q6b is highly related to two components, so should be eliminated

Rotated Component Matrix[a]

	Component		
	1	2	3
Q7a: Gov. responsibility: Provide job for everyone	.730		
Q7b: Gov. responsibility: Control prices	.691		
Q5e: Gov. and economy: Support declining industries to protect jobs	.681		
Q5f: Gov. and economy: Red. working week for more jobs	.533		
Q6c: Government should spend money: Law enforcement		.805	
Q6e: Government should spend money: Defence		.720	
Q6b: Government should spend money: Health		.453	.347
Q6a: Government should spend money: Environment			.831
Q6h: Government should spend money: Culture and arts			.740

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

a. Rotation converged in 4 iterations.

Now the rotated matrix looks very nice!

Rotated Component Matrix[a]

	Component		
	1	2	3
Q7a: Gov. responsibility: Provide job for everyone	.733		
Q7b: Gov. responsibility: Control prices	.697		
Q5e: Gov. and economy: Support declining industries to protect jobs	.686		
Q5f: Gov. and economy: Red. working week for more jobs	.535		
Q6a: Government should spend money: Environment		.831	
Q6h: Government should spend money: Culture and arts		.762	
Q6c: Government should spend money: Law enforcement			.790
Q6e: Government should spend money: Defence			.783

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

a. Rotation converged in 5 iterations.

Each factor explains a lot of the variance (at least 16%)

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	2.136	26.705	26.705	2.136	26.705	26.705	1.823	22.791	22.791
2	1.324	16.546	43.251	1.324	16.546	43.251	1.364	17.055	39.846
3	1.041	13.014	56.265	1.041	13.014	56.265	1.314	16.419	56.265
4	.905	11.316	67.581						
5	.763	9.537	77.118						
6	.675	8.439	85.558						
7	.595	7.437	92.995						
8	.560	7.005	100.000						

Extraction Method: Principal Component Analysis.

Now we can look at the test statistics

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.661
Bartlett's Test of Sphericity	Approx. Chi-Square	31305.627
	df	28
	Sig.	.000

- The Kaiser-Olkin measure is over.6, so it is acceptable
- And the Bartlett's Test is also significant
- So the model seems to be OK.

Now we must interpret these factors and give them names

- Factor 1 seems to do with the government intervening in the economy.
- Factor 2 seems to deal with government responsibility for the quality of life
- Factor 3 has to do with security.

Rotated Component Matrix[a]

	Component		
	1	2	3
Q7a: Gov. responsibility: Provide job for everyone	.733		
Q7b: Gov. responsibility: Control prices	.697		
Q5e: Gov. and economy: Support declining industries to protect jobs	.686		
Q5f: Gov. and economy: Red. working week for more jobs	.535		
Q6a: Government should spend money: Environment		.831	
Q6h: Government should spend money: Culture and arts		.762	
Q6c: Government should spend money: Law enforcement			.790
Q6e: Government should spend money: Defence			.783

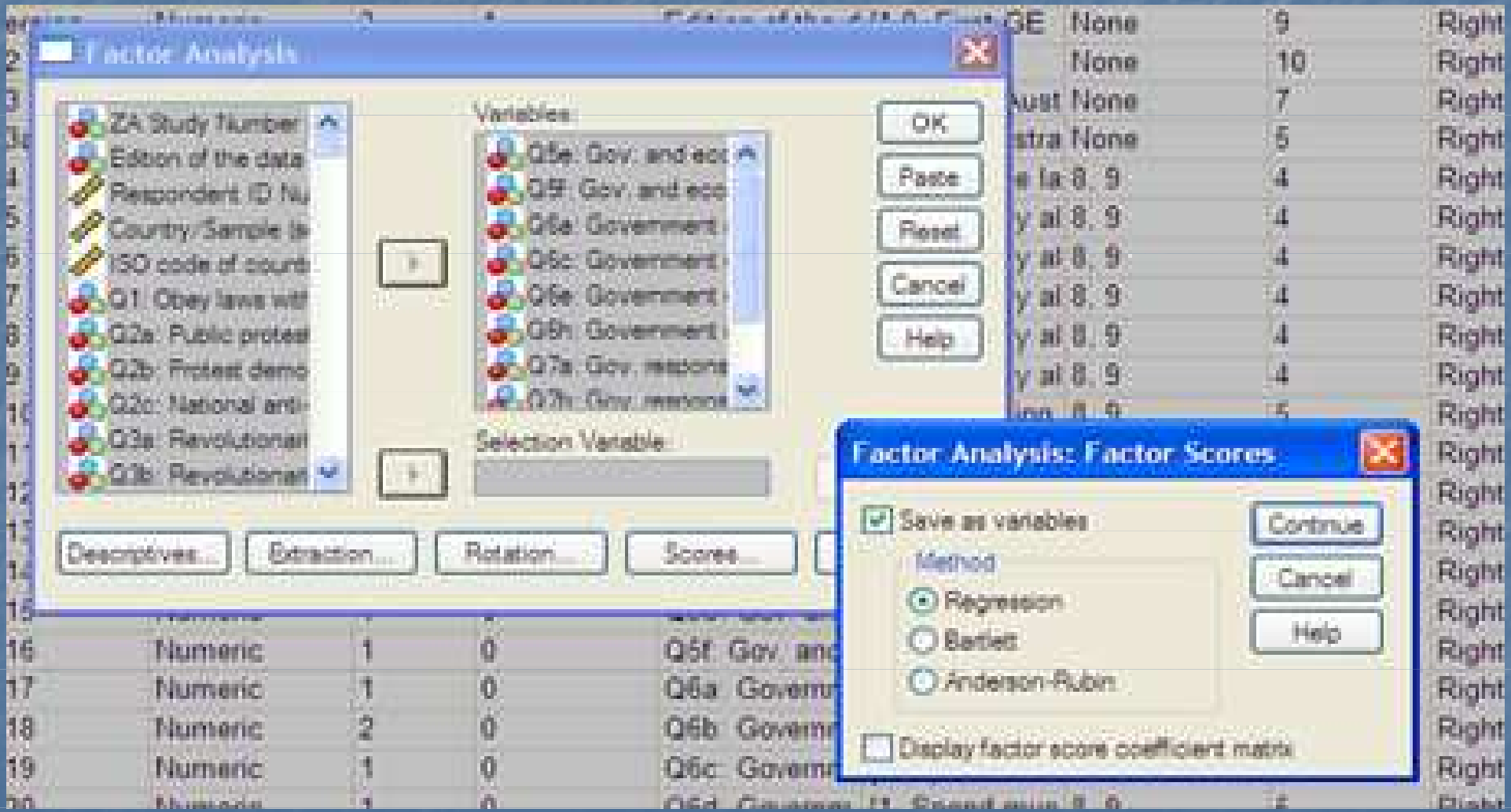
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Now there are two possibilities

- A: One can create three different scales by using the COMPUTE VARIABLE function as in creating the Cronbach Alpha scale
- Or one can let SPSS create a factor value for each factor.
- I favor the first method.
- To let SPSS create a factor value, go back to the function for factor analysis and click on SCORES…
- Then click on the save as variables box and press continue and run factor analysis one more time.

Creating factor scores



Now you can see that 3 new factors were created. You can rename them so that the factors have more meaning, such as GOVECINT

286	ve_size	Numeric	2	0	Size of community: Venezuela	{0, Not availabl	0	9	Right	Nominal
287	za_size	Numeric	2	0	Size of community: South Africa	{0, NAP, other	0	9	Right	Nominal
288	urbrural	Numeric	1	0	Type of community: Rs self-assessment	{0, Not availabl	0, 9	10	Right	Nominal
289	ethnic	Numeric	4	1	Family origin: ethnic group identity	{.0, Not availab	0, 99.0	8	Right	Scale
290	mode	Numeric	2	0	Administrative mode of data-collection	{10, F2f pap a	0	6	Right	Nominal
291	weight	Numeric	11	8	Weighting factor	{1.00000000,	None	13	Right	Scale
292	LESSREG	Numeric	8	2		{1.00, strongly	None	10	Right	Scale
293	Czech	Numeric	8	2		None	None	10	Right	Scale
294	filter_$	Numeric	1	0	Czech = 1 (FILTER)	{0, Not Selecte	None	10	Right	Scale
295	GOVREG	Numeric	8	2		None	None	10	Right	Scale
296	FAC1_1	Numeric	11	5	REGR factor score 1 for analysis 1	None	None	13	Right	Scale
297	FAC2_1	Numeric	11	5	REGR factor score 2 for analysis 1	None	None	13	Right	Scale
298	FAC3_1	Numeric	11	5	REGR factor score 3 for analysis 1	None	None	13	Right	Scale
299										

Finally you can conduct multivariate regressions on each factor

- Try to create a model for each 3 factor.
- Choose again some independent variables that you think might be able to explain attitudes toward these factors, such as AGE, SEX, EDUCATION, INCOME, etc.
- Examine whether the same independent variables are significant for each factor.
- If some variables are more important for explaining one factor than another, think about why this could be the case.

THEORY!

- Please remember that I have presented everything inductively now to show you how statistics work.
- You should actually begin with theory.
- From the beginning you should have hypotheses based on previous studies as to which variables should be able to predict your outcomes.
- Also you should make use of theory to choose what questions you will put in your original factor analysis. You should have a hypothesis about how many factors there will be.
- Even if your hypothesis is proven wrong, you should think theoretically about how to name the factors that you end up with.