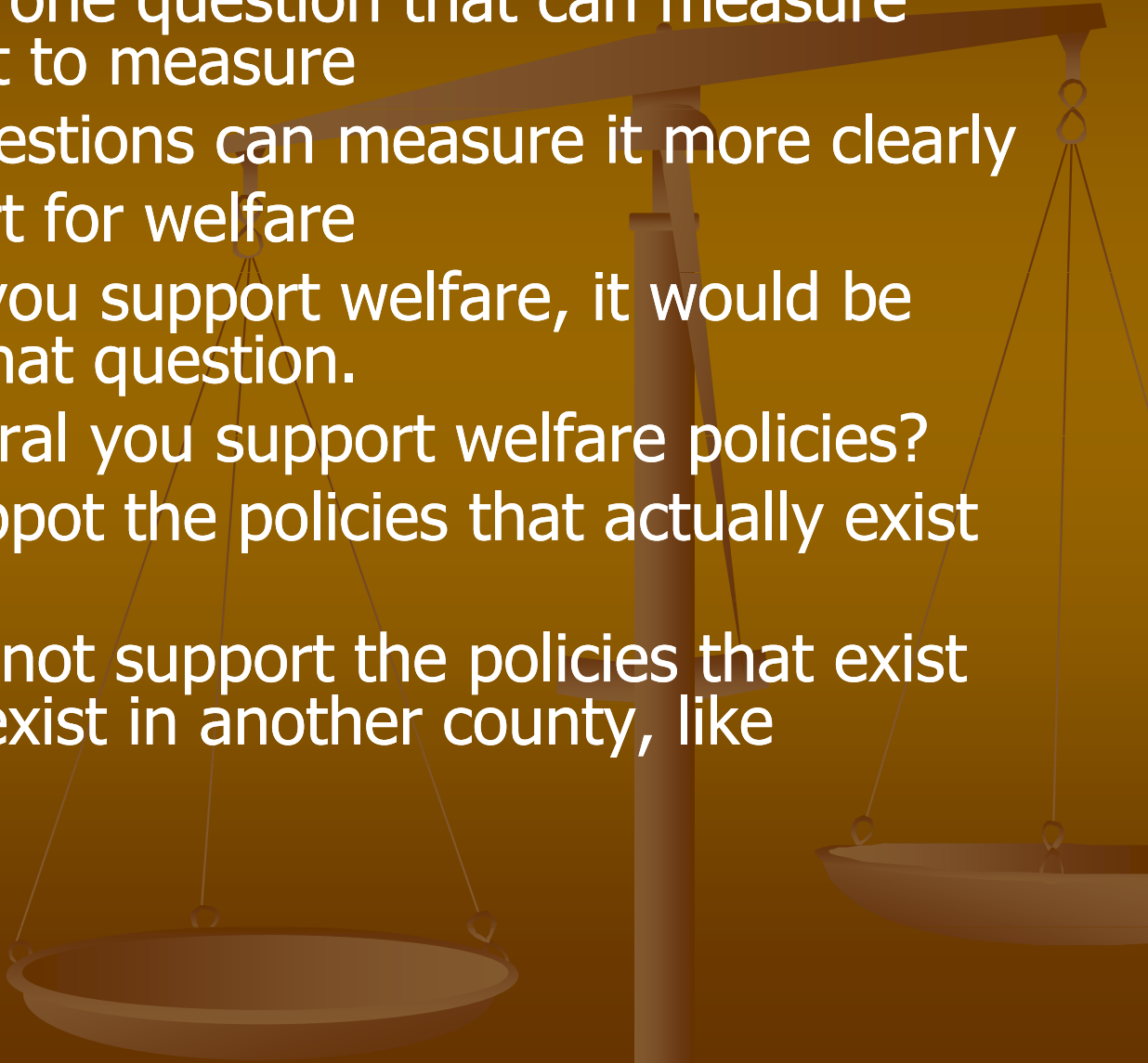# Scaling and Factor Analysis

Scaling: lack of a perfect question

- Often we cannot find one question that can measure exactly what we want to measure
- But a collection of questions can measure it more clearly
- An example is support for welfare
- If we would ask: do you support welfare, it would be difficult to interpret that question.
- Does it mean in general you support welfare policies?
- Does it mean you suppot the policies that actually exist in your own country?
- Does it mean you do not support the policies that exist in your country, but exist in another county, like Sweden?
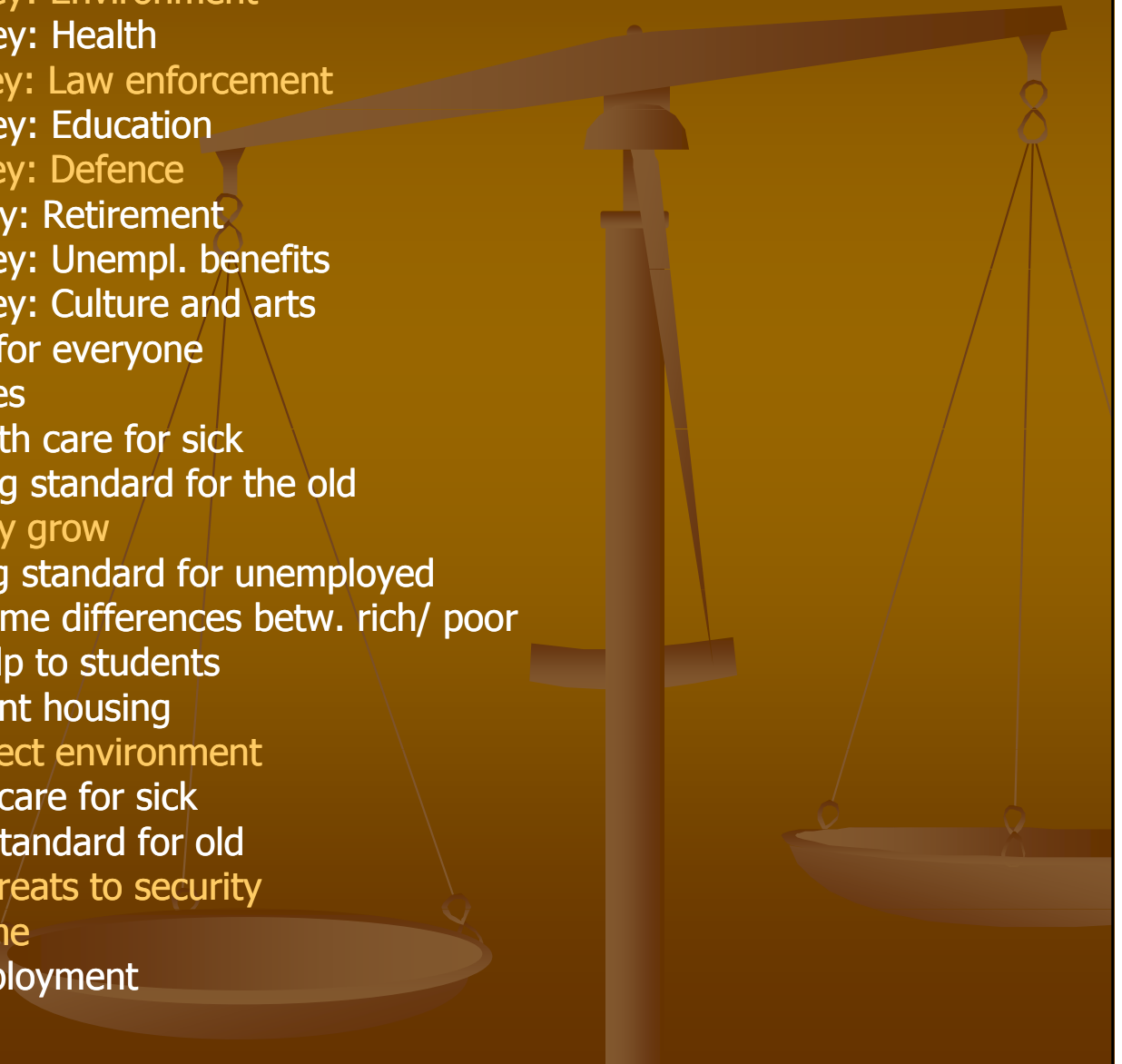
Questions from ISSP 2006 that deal with Welfare

- Q6a: Government should spend money: Environment
- Q6b: Government should spend money: Health
- Q6c: Government should spend money: Law enforcement
- Q6d: Government should spend money: Education
- Q6e: Government should spend money: Defence
- Q6f: Government should spend money: Retirement
- Q6g: Government should spend money: Unempl. benefits
- Q6h: Government should spend money: Culture and arts
- Q7a: Gov. responsibility: Provide job for everyone
- Q7b: Gov. responsibility: Control prices
- Q7c: Gov. responsibility: Provide health care for sick
- Q7d: Gov. responsibility: Provide living standard for the old
- Q7e: Gov. responsibility: Help industry grow
- Q7f: Gov. responsibility: Provide living standard for unemployed
- Q7g: Gov. responsibility: Reduce income differences betw. rich/ poor
- Q7h: Gov. responsibility: Financial help to students
- Q7i: Gov. responsibility: Provide decent housing
- Q7j: Gov. responsibility: Laws to protect environment
- Q8a: Gov. successful: Provide health care for sick
- Q8b: Gov. successful: Provide living standard for old
- Q8c: Gov. successful: Dealing with threats to security
- Q8d: Gov. successful: Controlling crime
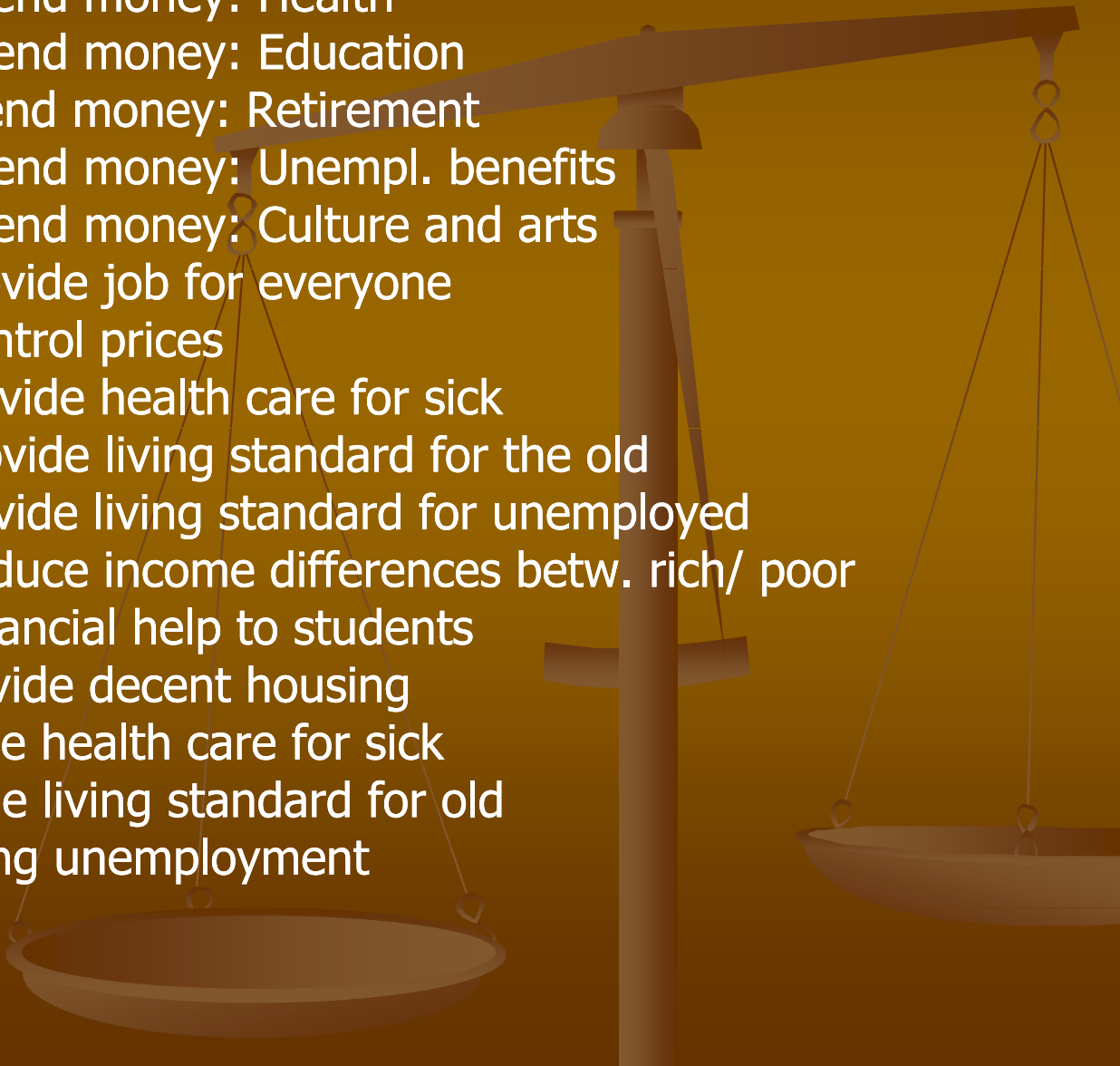- Q8e: Gov. successful: Fighting unemployment

The darker questions we could eliminate, because they do not deal with social policies

- Q6a: Government should spend money: Environment
- Q6b: Government should spend money: Health
- Q6c: Government should spend money: Law enforcement
- Q6d: Government should spend money: Education
- Q6e: Government should spend money: Defence
- Q6f: Government should spend money: Retirement
- Q6g: Government should spend money: Unempl. benefits
- Q6h: Government should spend money: Culture and arts
- Q7a: Gov. responsibility: Provide job for everyone
- Q7b: Gov. responsibility: Control prices
- Q7c: Gov. responsibility: Provide health care for sick
- Q7d: Gov. responsibility: Provide living standard for the old
- Q7e: Gov. responsibility: Help industry grow
- Q7f: Gov. responsibility: Provide living standard for unemployed
- Q7g: Gov. responsibility: Reduce income differences betw. rich/ poor
- Q7h: Gov. responsibility: Financial help to students
- Q7i: Gov. responsibility: Provide decent housing
- Q7j: Gov. responsibility: Laws to protect environment
- Q8a: Gov. successful: Provide health care for sick
- Q8b: Gov. successful: Provide living standard for old
- Q8c: Gov. successful: Dealing with threats to security
- Q8d: Gov. successful: Controlling crime
- Q8e: Gov. successful: Fighting unemployment

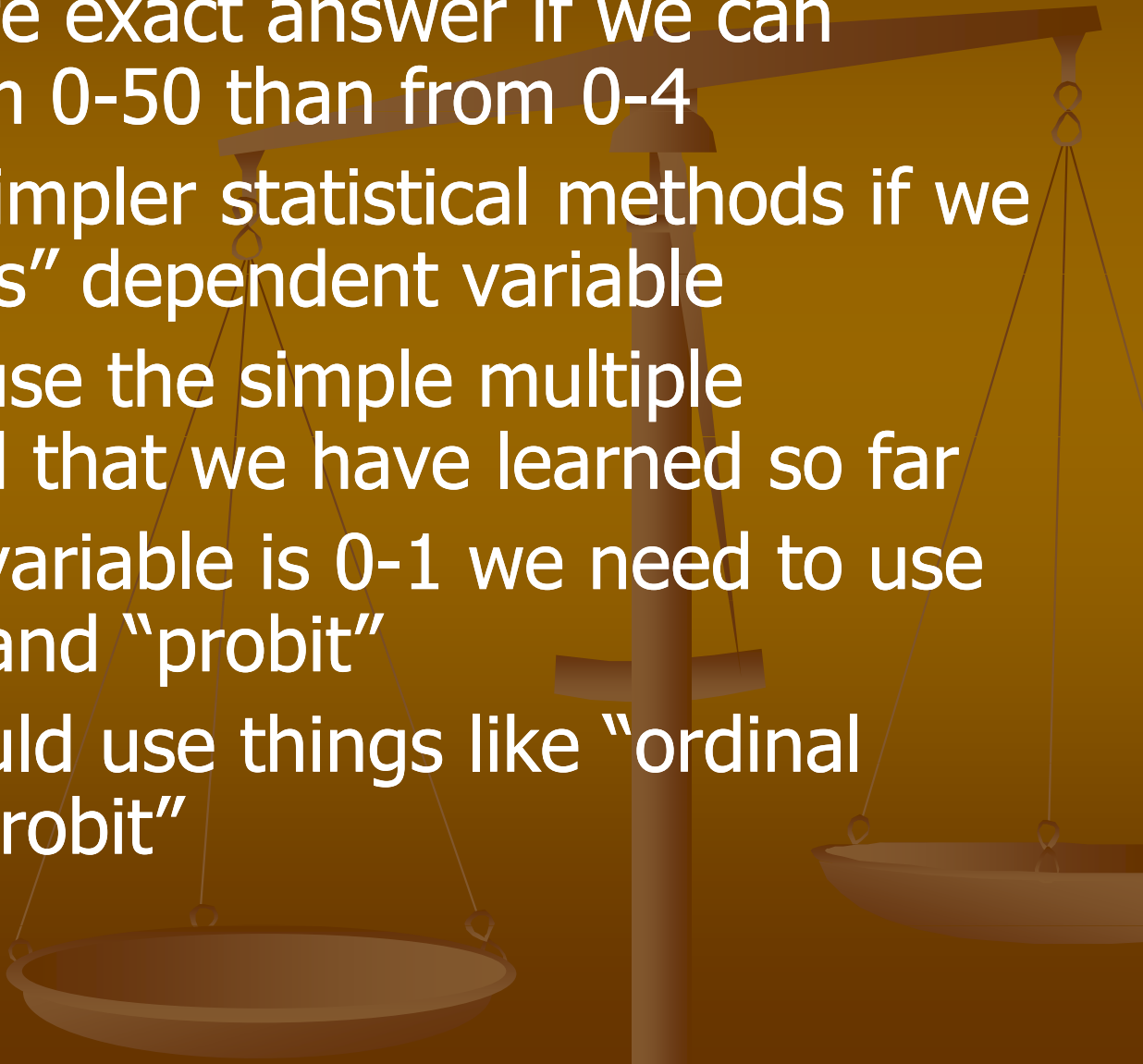# After Eliminating the Darker ones, we see 3 groups: 1)spending money, 2) responsibility, 3) successful

- Q6b: Government should spend money: Health
- Q6d: Government should spend money: Education
- Q6f: Government should spend money: Retirement
- Q6g: Government should spend money: Unempl. benefits
- Q6h: Government should spend money: Culture and arts
- Q7a: Gov. responsibility: Provide job for everyone
- Q7b: Gov. responsibility: Control prices
- Q7c: Gov. responsibility: Provide health care for sick
- Q7d: Gov. responsibility: Provide living standard for the old
- Q7f: Gov. responsibility: Provide living standard for unemployed
- Q7g: Gov. responsibility: Reduce income differences betw. rich/ poor
- Q7h: Gov. responsibility: Financial help to students
- Q7i: Gov. responsibility: Provide decent housing
- Q8a: Gov. successful: Provide health care for sick
- Q8b: Gov. successful: Provide living standard for old
- Q8e: Gov. successful: Fighting unemployment

Implications

- The ones about success are new and to make it simple, I will ignore them for now

- Spending: problem is that one's answer depends on one's starting point. A Swede who thinks the government should spend less might be more favorable to the welfare state than an American who thinks the government should spend more, because they have different starting points

- Responsibility: one can think the government should be resonsible for healthcare AND provide it, or simply that it should be responsible for healthcare by REGULATING it and keeping it private

- Since no perfect questions exist, we get a better idea by combining imperfect questions

The advantage in have a larger number of possible outcomes

- We also get a more exact answer if we can create a scale from 0-50 than from 0-4
- We can also use simpler statistical methods if we have a "continuous" dependent variable
- This allows us to use the simple multiple regression method that we have learned so far
- If the dependent variable is 0-1 we need to use things like "logit" and "probit"
- If it is 0-4 we should use things like "ordinal logit" or "ordinal probit"

One Dimensional Scales

- Cronbach alpha
- Test to see if all the questions are consistent
- We might think a group of questions belong together, but the respondents could interpret them differently
- Cronbach's alfa expects all the questions to measure approximately the same thing
- But let´s say the scale is 0-4. Perhaps my "true" score is 2.8 and your true score is 2.7. If we only have one question, then we will both answer 3, so more questions makes it more accurate, as my average then would become 2.8 and yours 2.7.
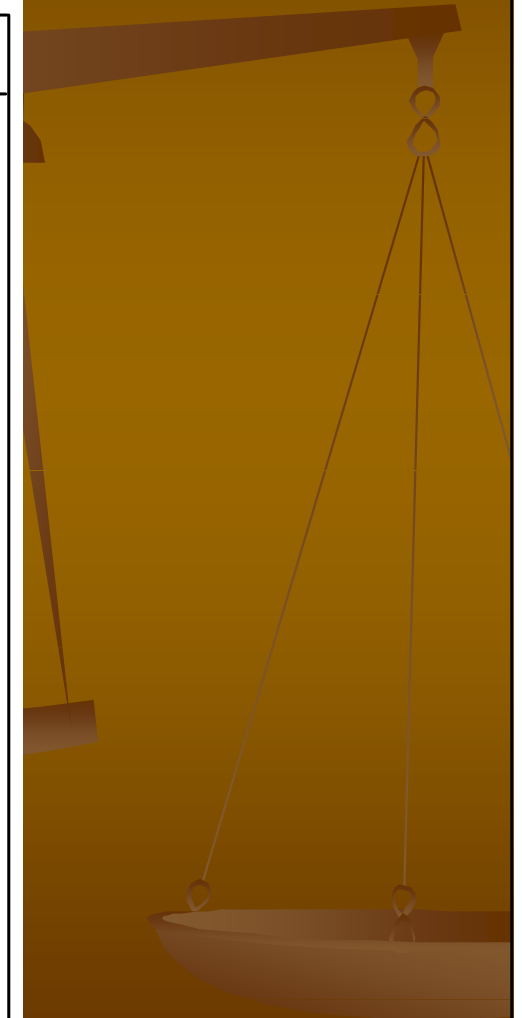- It is acceptable if alpha >.6 but best if >.8

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
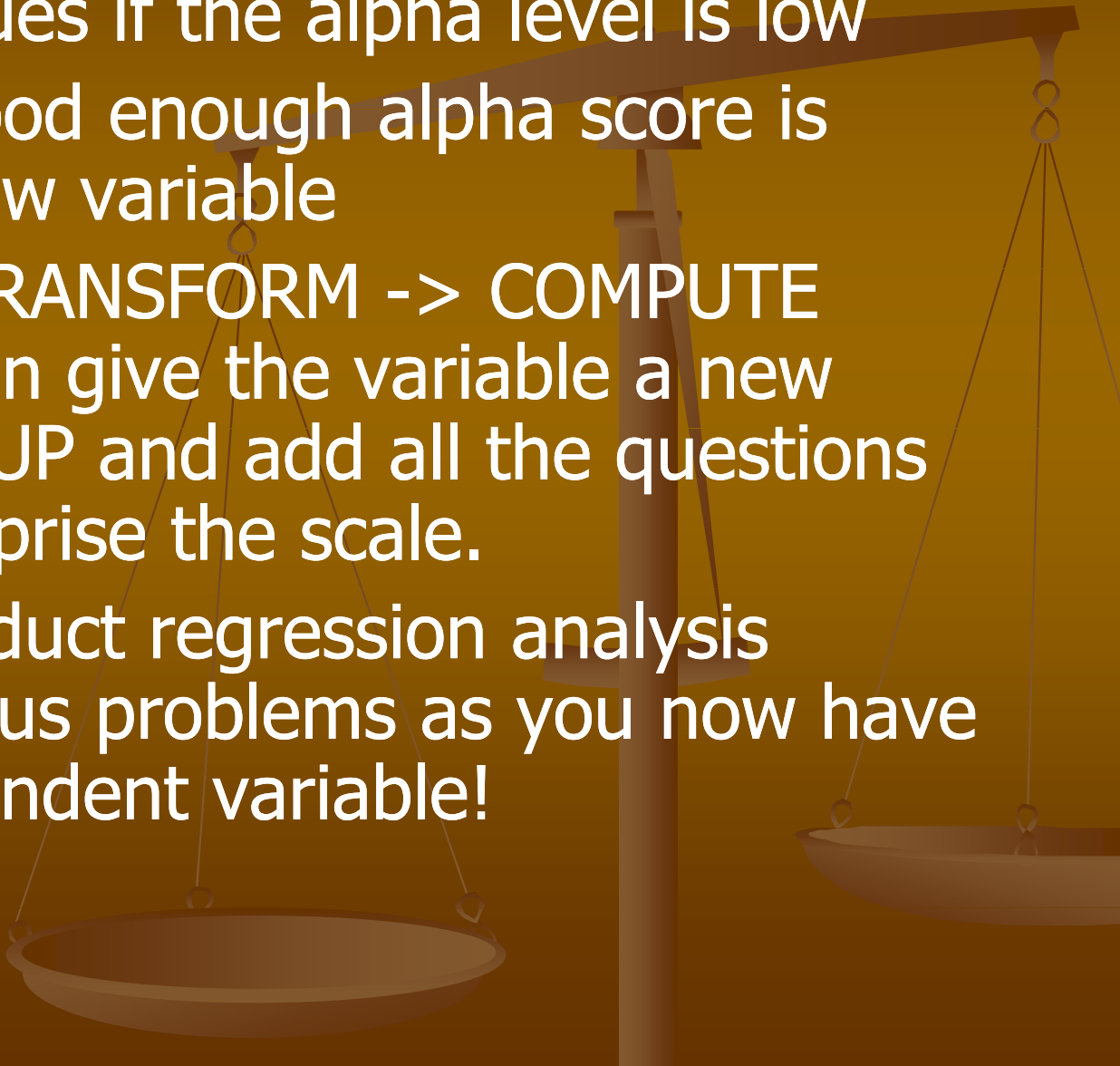
Reliability Statistics

Cronbach's Alpha	N of Items
.675	16

We look at the last row and see which items would increase the alpha score if eliminated
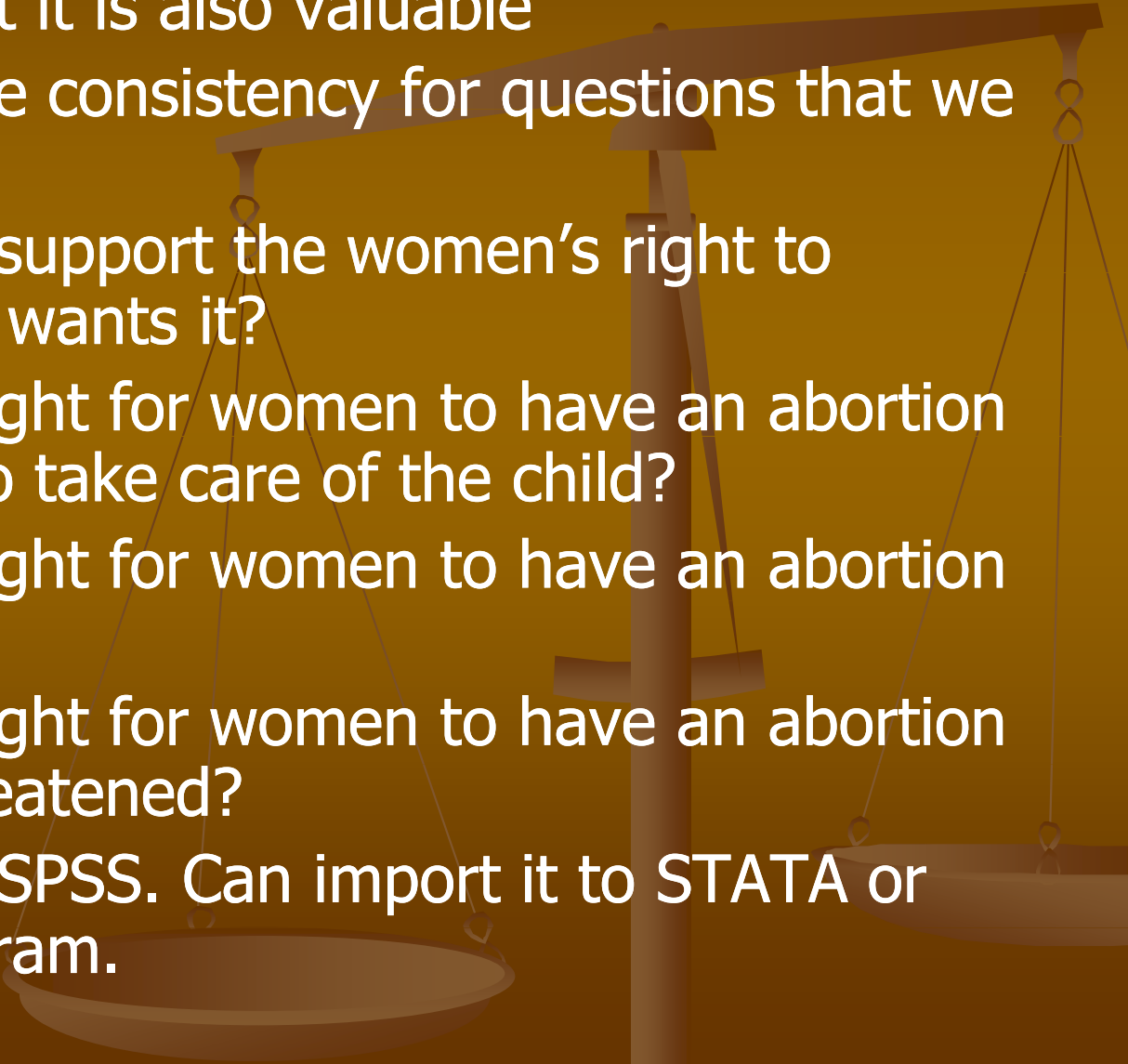
Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Q5a: Gov. and economy: Cuts in gov. spending	38.48	38.062	-.037	.067	.704
Q5b: Gov. and economy: Financing projects for new jobs	38.79	33.260	.417	.257	.643
Q5c: Gov. and economy: Less gov. reg. of business	38.51	36.464	.103	.146	.684
Q5d: Gov. and economy: Support industry to develop new products	38.58	33.468	.376	.272	.648
Q5e: Gov. and economy: Support declining industries to protect jobs	37.87	31.666	.431	.320	.637
Q5f: Gov. and economy: Red. working week for more jobs	37.71	34.531	.226	.141	.669
Q6a: Government should spend money: Environment	38.33	34.922	.328	.239	.656
Q6b: Government should spend money: Health	38.76	34.806	.345	.224	.654
Q6c: Government should spend money: Law enforcement	38.04	34.856	.305	.218	.658
Q6d: Government should spend money: Education	38.66	35.122	.305	.327	.658
Q6e: Government should spend money: Defence	37.39	35.746	.210	.156	.669
Q6f: Government should spend money: Retirement	38.56	34.101	.389	.246	.648
Q6g: Government should spend money: Unempl. benefits	37.56	34.475	.306	.214	.657
Q6h: Government should spend money: Culture and arts	37.78	35.236	.300	.221	.659
Q7a: Gov. responsibility: Provide job for everyone	38.94	34.990	.268	.293	.662
Q7b: Gov. responsibility: Control prices	38.60	35.019	.265	.269	.662

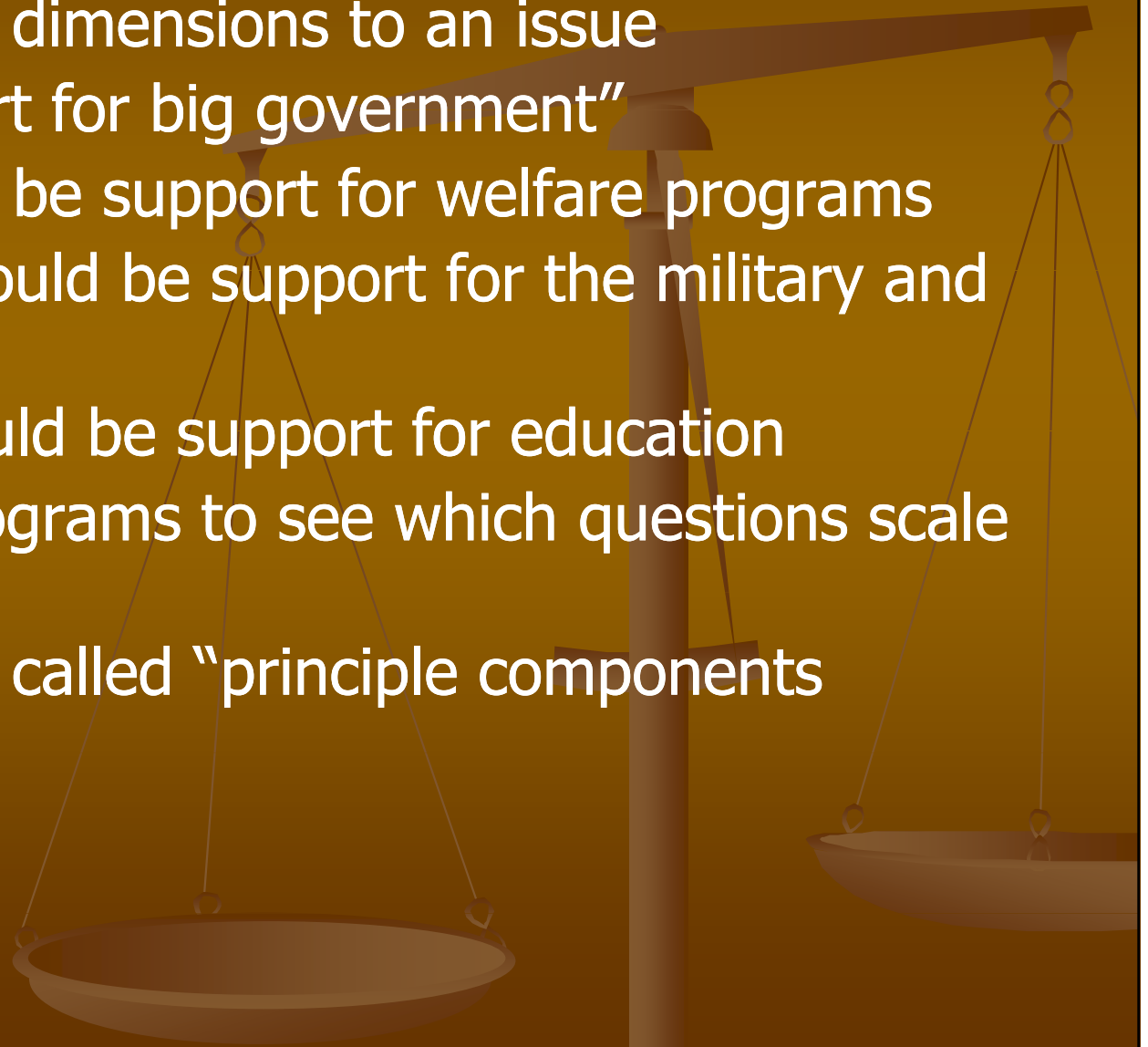# What to do after conducting reliability analysis?

- First eliminate values if the alpha level is low
- Second when a good enough alpha score is found, create a new variable
- You again go to TRANSFORM -> COMPUTE VARIABLE and then give the variable a new name, like WELFSUP and add all the questions together that comprise the scale.
- Then you can conduct regression analysis without the previous problems as you now have a continuous dependent variable!
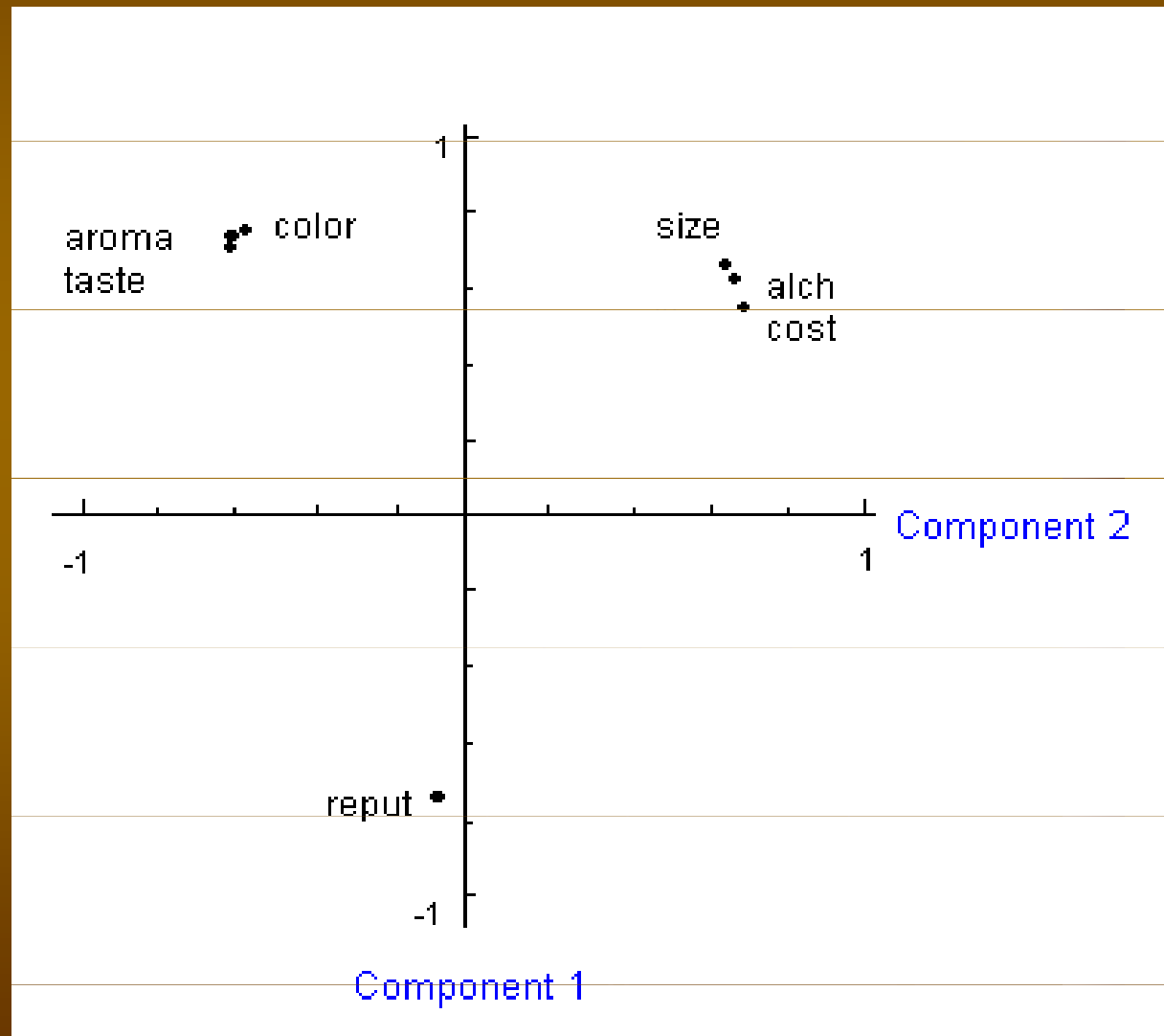
Mokken Scale

- Is not used much, but it is also valuable
- It allows us to test the consistency for questions that we can rank
- For example: do you support the women's right to abortion anytime she wants it?
- Do you support the right for women to have an abortion if they are too poor to take care of the child?
- Do you support the right for women to have an abortion if they were raped?
- Do you support the right for women to have an abortion if their own life is threatened?
- Unfortunately, not in SPSS. Can import it to STATA or buy as separate program.
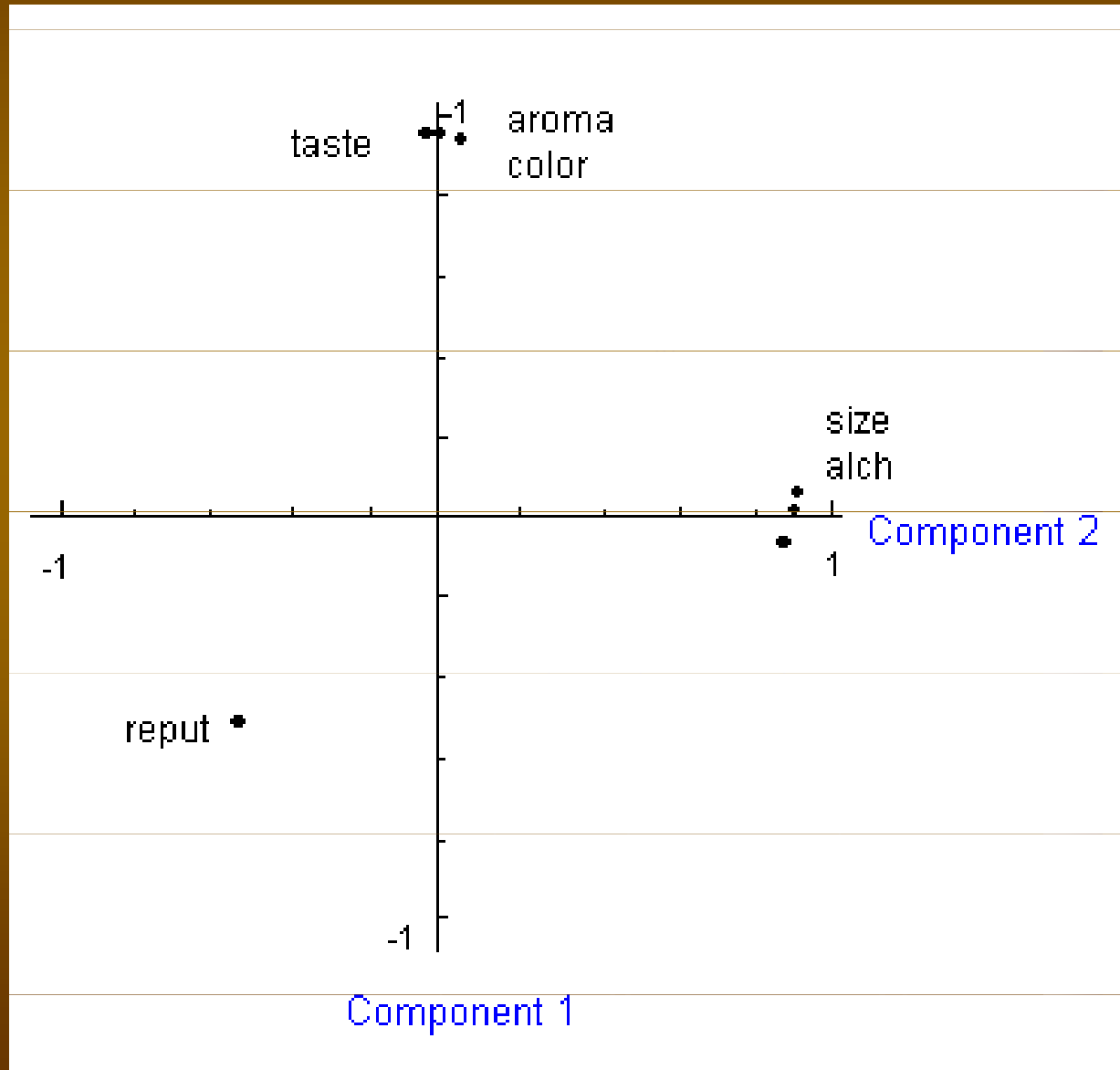
Multidimensional scaling

- Factor analysis

- There can be several dimensions to an issue

- For example: "support for big government"

- One dimension could be support for welfare programs

- Another dimension could be support for the military and police

- A third dimension could be support for education

- We use statistical programs to see which questions scale well together.
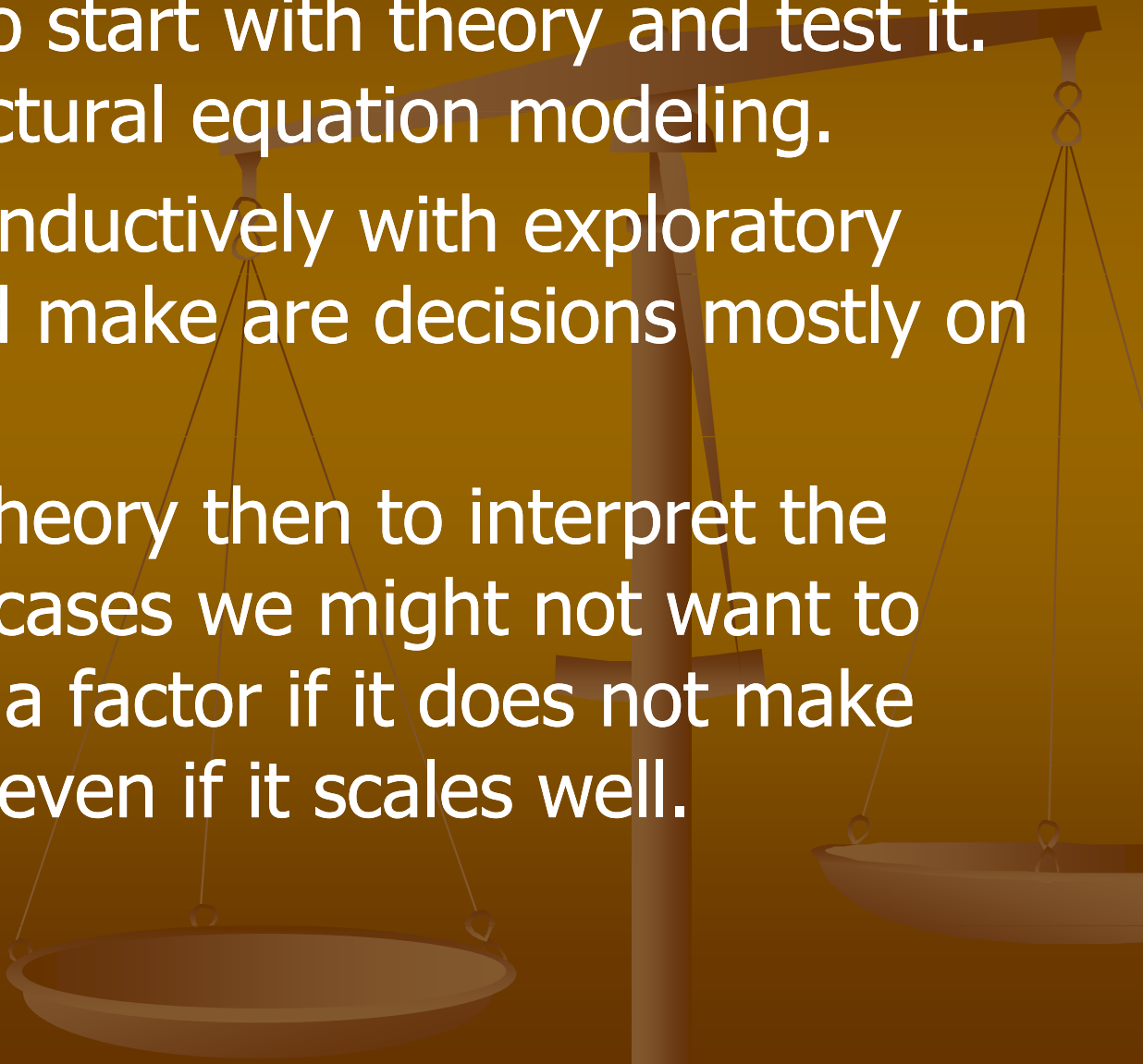
- The most common is called "principle components analysis"

Rotation: we don´t want the items in the two factors to be related, so we rotate the axis: before rotation

After Rotation (REPUT should be revmoved as it is related strongly to both factors)

How Many Factors are There?

- It would be best to start with theory and test it. We do this in structural equation modeling.

- In SPSS we work inductively with exploratory factor analysis and make are decisions mostly on data.

- But we still need theory then to interpret the data and in some cases we might not want to include an item in a factor if it does not make theoretical sense, even if it scales well.
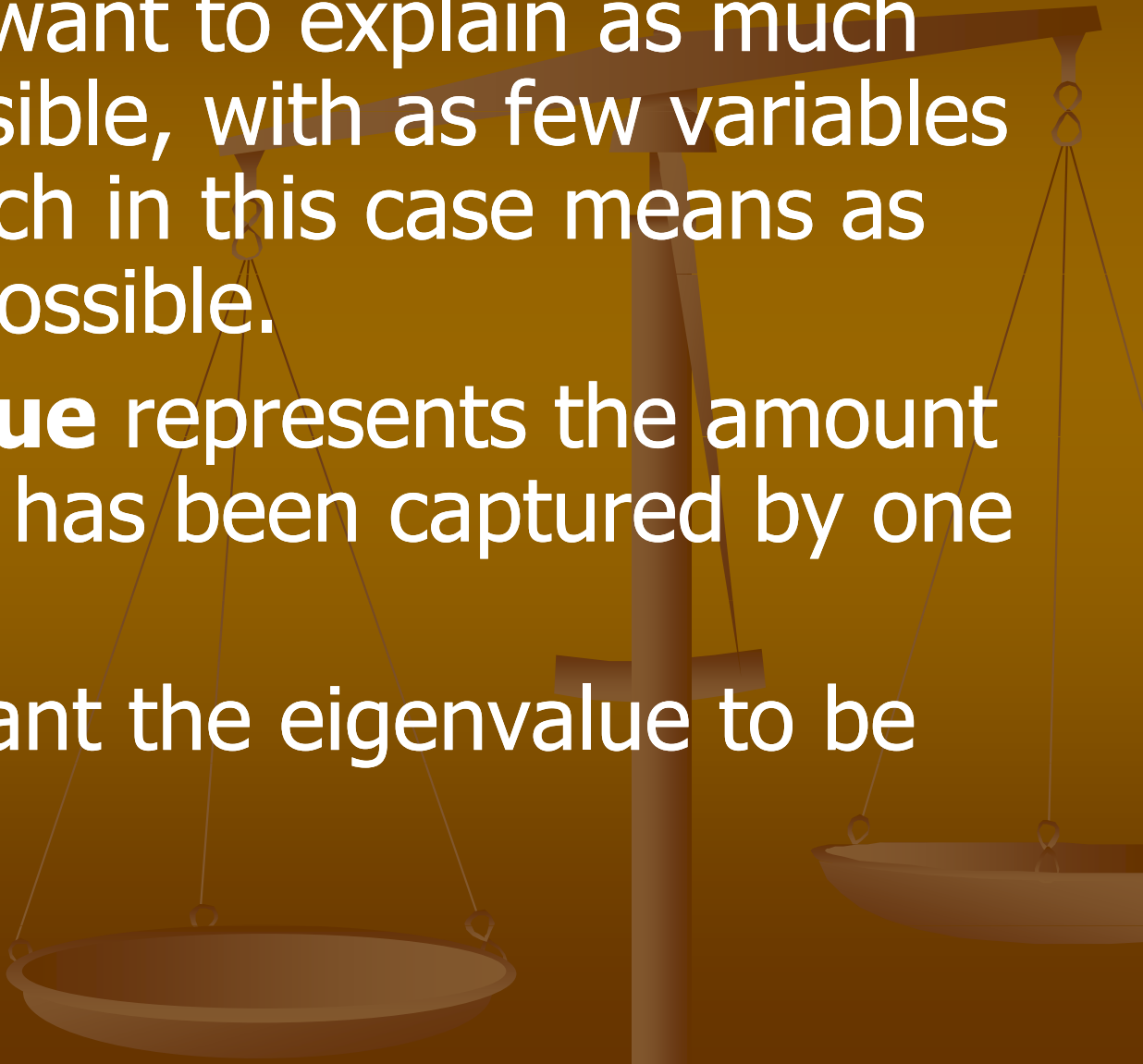
Explained variance

- As is usual, we want to explain as much variance as possible, with as few variables as possible, which in this case means as few factors as possible.

- Each **eigenvalue** represents the amount of variance that has been captured by one component.

- Normally, we want the eigenvalue to be greater than 1.

Table of Eigenvalues

| Component | Initial Eigenvalues | | |
	Total	% of Variance	Cumulative %
1	3.313	47.327	47.327
2	2.616	37.369	84.696
3	.575	8.209	92.905
4	.240	3.427	96.332
5	.134	1.921	98.252
6	9.E-02	1.221	99.473
7	4.E-02	.527	100.000

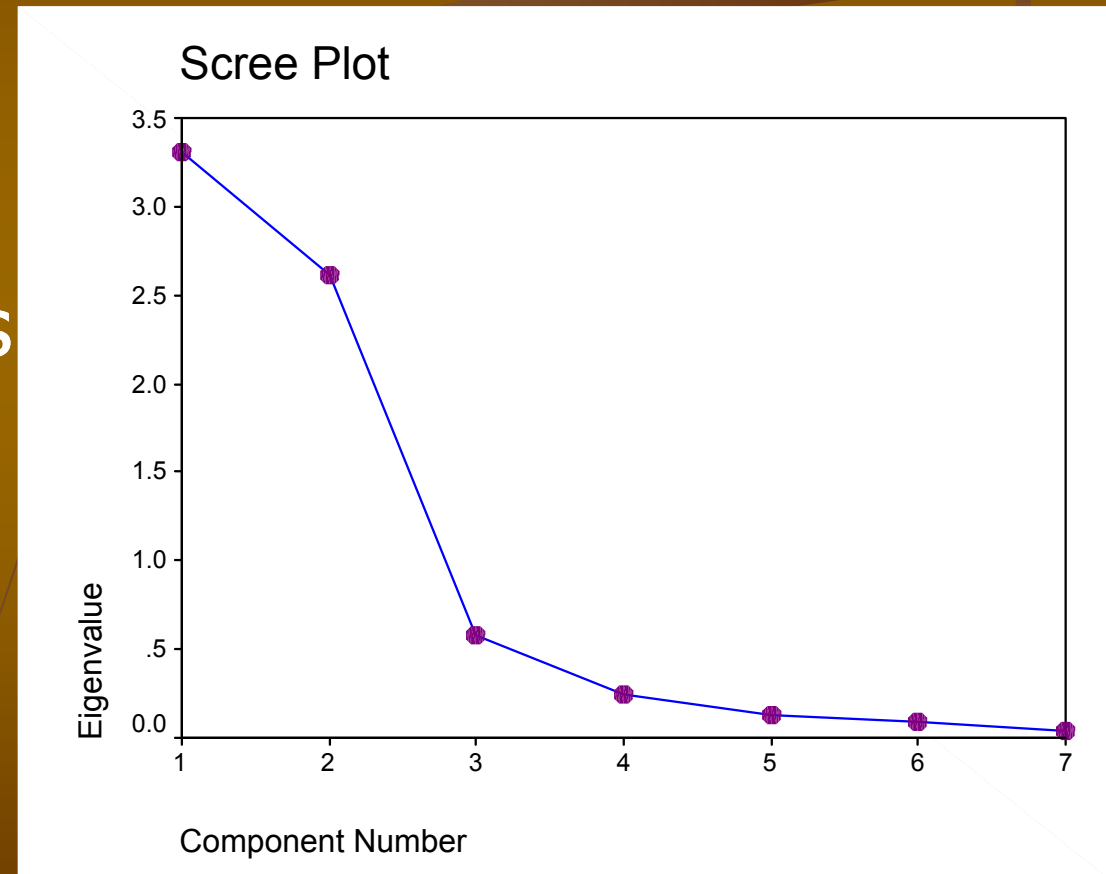Extraction Method: Principal Component Analysis.

- The first two components have eigenvalues greater than 1.

- Their account for almost 85% of the explained variance.

- Addint a third factor would only increase the explained variance by a little more than 8%.

Scree Plot

- We look where the biggest drop takes place.

- We see that after the second factor, there is a huge drop in Eigenvalue.

- So again, it seems there are only two factors.

Scree Plot

Testing

- Kaiser has described MSAs above .9 as marvelous, above .8 as meritorious, above .7 as middling, above .6 as mediocre, above .5 as miserable, and below .5 as unacceptable.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.665
Bartlett's Test of Sphericity	Approx. Chi-Square	1637.9
	df	21
	Sig.	.000

After creating factors

- SPSS lets you create new variables for each factor.
- This is fine if you only want to regressions on each factor.
- If you want to compare outcomes for each factor and be able to say, for example, that Swedes score higher than Czechs on Factor 1 (support for equality) then it is good to make your own new variables by adding together all the items (i.e. questions) for each factor.
- In other words, you do just as with Cronbach´s alpha, but this time you create several scales, not just one.
- Afterwards, you can run regressions on each factor separately just as with Cronbach´s alpha.
- In advanced statistics, such as Structural Equation Modelling, it is also possible to include sevaral factors in one model and run several regressions simultaneously, but this is not so common even if I usually do it this way.