

# Workshop Construction and Analysis of Tests and Questionnaires

Day 2: Validity (Factor analysis)

# Simulated data for illustrative purposes

Generated/fictitious dataset with six items X1 upto and including X6  
Depression-scale

N respondents = 300

*Correlation matrix:*

**Correlation Matrix**

		Anxious	Tense	Restless	Depressed	Useless	Unhappy
Correlation	Anxious	1,000	,449	,443	,296	,314	,326
	Tense	,449	1,000	,446	,312	,264	,250
	Restless	,443	,446	1,000	,279	,258	,282
	Depressed	,296	,312	,279	1,000	,467	,516
	Useless	,314	,264	,258	,467	1,000	,497
	Unhappy	,326	,250	,282	,516	,497	1,000

**Correlation Matrix**

		Anxious	Tense	Restless	Depressed	Useless	Unhappy
Correlation	Anxious	1,000	,449	,443	,296	,314	,326
	Tense	,449	1,000	,446	,312	,264	,250
	Restless	,443	,446	1,000	,279	,258	,282
	Depressed	,296	,312	,279	1,000	,467	,516
	Useless	,314	,264	,258	,467	1,000	,497
	Unhappy	,326	,250	,282	,516	,497	1,000

Can you see a pattern? Are there multiple dimensions?

**Correlation Matrix**

		X1: Anxious	X2: Tense	X3: Restless	X4: Depressed	X5: Useless	X6: Unhappy
Correlation	X1: Anxious	1,000	,449	,443	,296	,314	,326
	X2: Tense	<u>,449</u>	1,000	,446	,312	,264	,250
	X3: Restless	<u>,443</u>	<u>,446</u>	1,000	,279	,258	,282
	X4: Depressed	,296	<u>,312</u>	,279	1,000	,467	,516
	X5: Useless	<u>,314</u>	,264	,258	<u>,467</u>	1,000	,497
	X6: Unhappy	<u>,326</u>	,250	,282	<u>,516</u>	<u>,497</u>	1,000

- For now: we just assume that we have two factors.  
*We will learn later how we can determine the number of factors*
- Two factors -> 2 dimensions -> 2 scales
  - Factor 1: Accounts for most of the variance seen in the six items
  - Factor 2: **given the first factor and in addition to this factor** accounts for most of the *remaining* variance of the six items

# Factor loadings

Factor Matrix <sup>a</sup>

	Factor	
	1	2
X1: Anxious	,610	,274
X2: Tense	,583	,345
X3: Restless	,575	,327
X4: Depressed	,642	-,261
X5: Useless	,616	-,268
X6: Unhappy	,665	-,337

Extraction Method: Principal Axis Factoring.

a. 2 factors extracted. 10 iterations required.

- Here, in this table we see the factorloadings for every item on factor 1 and factor 2
- Note that the second factor accounts for unexplained var/correlation *on top* of the 1st component
- The **loadings** on the first and second factor are equal to the correlation of an item with these factors (dimensions):
  - $a_{j1} = r_{X_j F_1}$
  - $a_{j2} = r_{X_j F_2}$
- For example: the correlation of the item 'Anxious' with the second factor = 0,375:
  - $r_{Anxious, F_2} = a_{12} = 0,375$

# Communality

Factor Matrix <sup>a</sup>

	Factor	
	1	2
X1: Anxious	,610	,274
X2: Tense	,583	,345
X3: Restless	,575	,327
X4: Depressed	,642	-,261
X5: Useless	,616	-,268
X6: Unhappy	,665	-,337

Extraction Method: Principal Axis Factoring.

a. 2 factors extracted. 10 iterations required.

- The **communality**  $h_j^2$  of  $X_j$  is the amount of variance of  $X_j$  accounted for by *both* factors
- $h_j^2 = r_{X_j F_1}^2 + r_{X_j F_2}^2 = a_{j1}^2 + a_{j2}^2$
- For example: the factors account for 0,447 of the variance in the scores on X1: 'Anxious'
- $h_1^2 = a_{11}^2 + a_{12}^2 = (0,610)^2 + (0,274)^2 = 0,447$

# Variance accounted for ('eigen value'):

Factor Matrix <sup>a</sup>

	Factor	
	1	2
X1: Anxious	,610	,274
X2: Tense	,583	,345
X3: Restless	,575	,327
X4: Depressed	,642	-,261
X5: Useless	,616	-,268
X6: Unhappy	,665	-,337

Extraction Method: Principal Axis Factoring.

a. 2 factors extracted. 10 iterations required.

- The **variance accounted for by a factor**

- $VAF_2 = a_{12}^2 + a_{22}^2 + a_{32}^2 + a_{42}^2 + a_{52}^2 + a_{62}^2$

- For example: the **second** factor accounts for 0,554 variance in total

- $VAF_2 = 0,554$



# And in SPSS:

## Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,802	46,700	46,700	2,276	37,937	37,937
2	1,081	18,015	64,715	,554	9,228	47,165
3	,578	9,639	74,354			
4	,560	9,335	83,690			
5	,522	8,693	92,382			
6	,457	7,618	100,000			

Extraction Method: Principal Axis Factoring.

Total variance accounted for by PAF

- So, how do we know how many factors we need?!
- Are there any rules to decide on the number of factors?

# How do we decide how many factors we have?

- How can you determine the number of factors underlying the data?
  - Option 1: Common sense
  - Option 2: Rules based on eigenvalues: Kaiser's criterion and scree test
  - Option 3: Theory

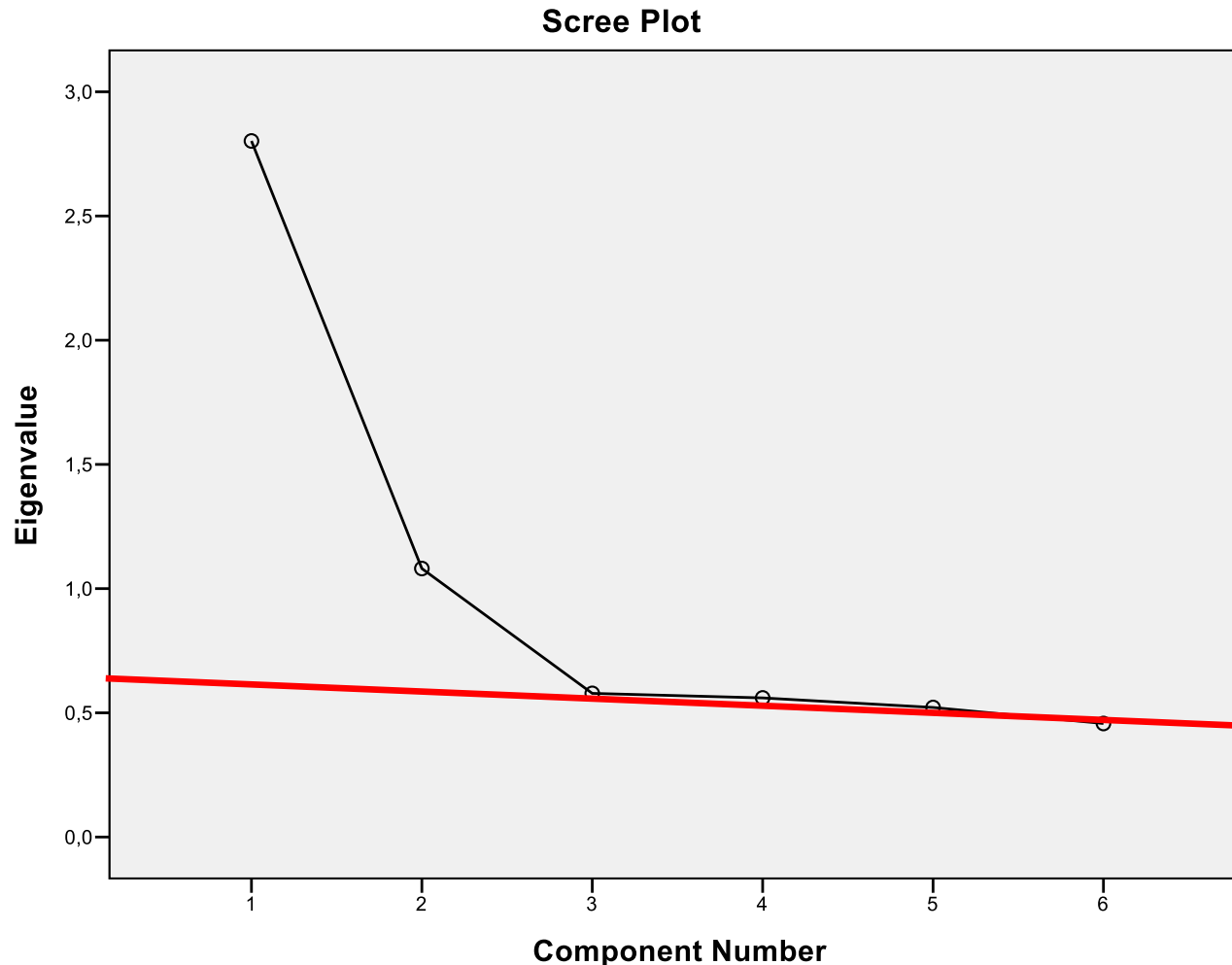
# (1) Common sense

- $K \leq J/3$
- Intuition: a factor with less than three items cannot be reliable
- For the example:  $K \leq J/3 = 6/3 = 2$   
( $K$  = total number of factors,  $J$  = total number of items)

## (2) Kaiser's criterion

- “Greater-than-one” rule
- Number of factors = number of factors having an eigenvalue larger than one
- How many factors would we choose for our example, based on this rule?

### (3) Scree test (Catell)



- Choose the number of factors for which the eigenvalue is above the line that goes true the “scree”
- Often very subjective
- Here: indicates a solution with 2 factors

## (4) Theory

- Often there is some theoretical knowledge about the number of constructs/dimensions underlying the data
- Eg, NEO big five
- Use this knowledge to further support other indications of that number of factors or to choose between several options (in case one rule or different rules do not result in a unique number)
  - eg, scree test: often 'different levels of scree', or different number of factors based on Kaiser's criterion and the scree plot

# Methods to find easier solutions for a factor analysis



# Loading: what dimension does an item measure?

- Results of PCA and PAF: can be used to determine, per item, to which dimension this item “belong” (based on component loadings / factor loadings)
- So: to determine the internal structure: Which loading is the highest? Highest loading indicates to which this item belongs
- However, this is not always as easy as it sounds..

# SPSS output in the perfect world

- What would be ideal would be that we don't have any doubts to which dimension an item belongs to
- E.g., perfect correlation/loading 1 on only ONE dimension + rest of the loadings = 0
- Immediately clear to what dimension an item belongs
- => *easy interpretation!*
- *But in practice, we don't get that (to the right: hypothetical example!)*

Component Matrix <sup>a</sup>			
	Component		
	Component 1	Component 2	Component 3
Item 1	1	0	0
Item 2	1	0	0
Item 3	1	0	0
Item 4	0	1	0
Item 5	0	1	0
Item 6	0	1	0
Item 7	0	0	1
Item 8	0	0	1
Item 9	0	0	1
Extraction Method: Principal Component Analysis.			

... in real life: interpretation of the output can be difficult:

Component Matrix <sup>a</sup>			
	Component		
	1	2	3
+ situatie denk	.837	-.258	
- situatie denk	.819	-.245	
- denk	.762	-.230	
+ denk	.744	-.260	-.106
verkeer	.687		
controleer	.370	.761	
voor mezelf	.265	.745	.133
- uitdrukking	.279	.658	.324
+ uitdrukking	.253	.413	-.721
stressvol	.373		.580
Extraction Method: Principal Component Analysis.			
a. 3 components extracted.			
b. Values < 0.2 have been left blank (not shown in the table)			

**Component Matrix <sup>a</sup>**

	Component	
	1	2
X1: Anxious	.687	.375
X2: Tense	.654	.467
X3: Restless	.651	.464
X4: Depressed	.708	-.385
X5: Useless	.688	-.415
X6: Unhappy	.710	-.432

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

**Factor Matrix**

	Factor		
	1	2	3
ACT1	.321	-.221	.689
ACT2	.587	.103	.132
ACT3	.623	.589	-.536
ACT4	.214	-8.75E-02	.379
ACT5	.709	.563	-.221
ACT6	-.240	.426	4.446E-02
ACT7	.549	-.191	.361
ACT8	.327	.307	1.874E-02
ACT9	-.215	.380	.144

Extraction Method: Principal Axis Factoring.

- “Perfect” loadings of 0/1 => unfortunately, utopic..
- In practice: Realistic aim:
  - as simple as possible (-> easier interpretation!)
  - We try to get as close as possible to a “simple structure” :
- **Rules of thumb for a simple structure**
  - Every item (variable) measures only ONE subconstruct (dimension) = **every item has a loading of >|,30| with only one factor**

# How do we get close(r) to a simple structure?

- The solution we get from a factor analysis is an *estimation* and therefore **not unique**:
  - There are endless possible combinations of factorloadings that are all equally good
    - the different solutions explain the same amount of variance & reproduce the same residual correlations
    - Compare with: different regression equations that explain the same amount of variance
  - Simply put:  $5 + 5 = 10$ , but  $4 + 6$  also equals 10.

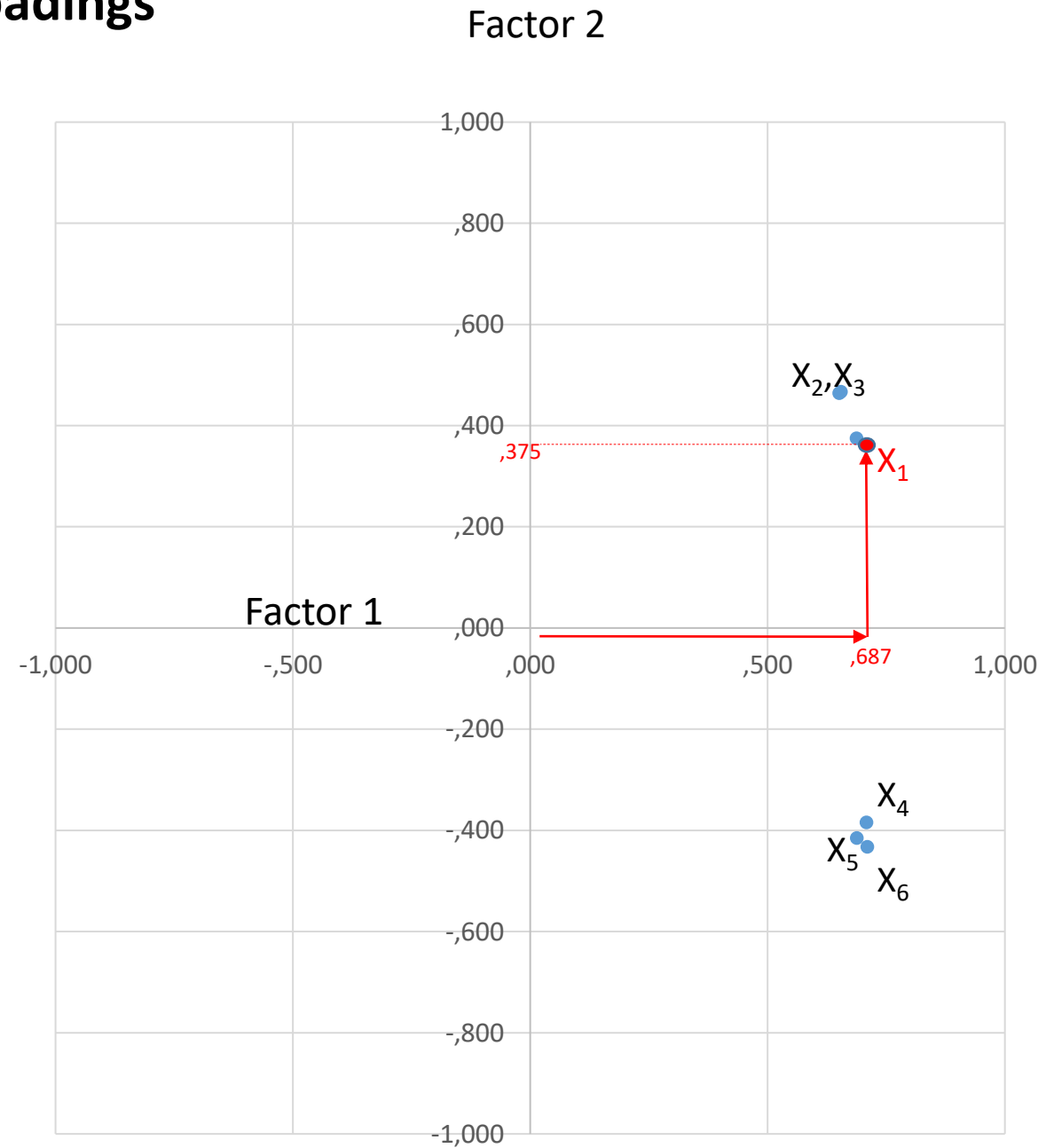
# Get closer to a simple structure: ROTATION

- So we are looking for an *alternative combination* of factorloadings that
  - Are of the same quality (explain an equal amount of variance)
  - But: are easier to interpret

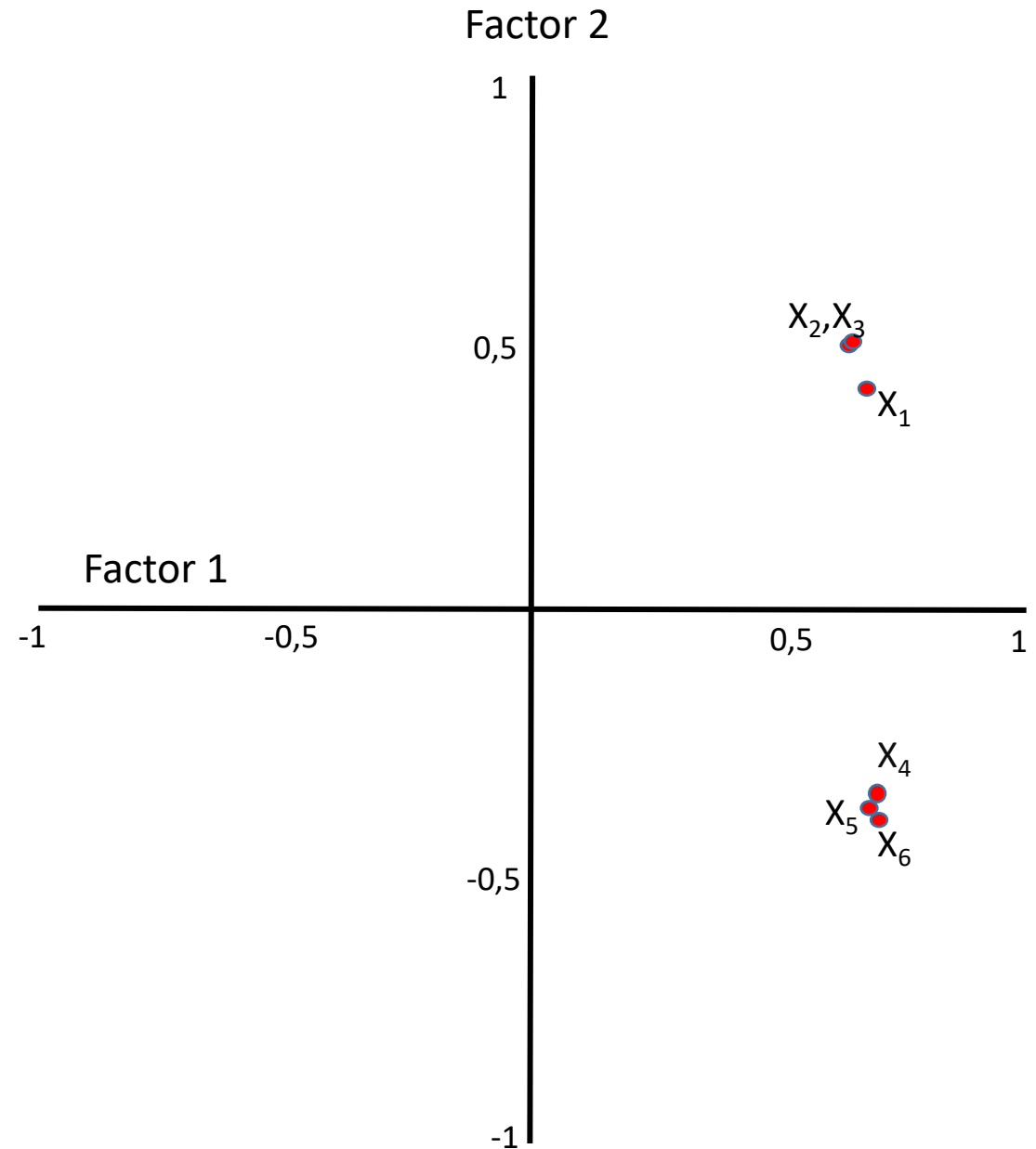
How?? -> Statistical trick 😊: ROTATION

## Figure of the loadings

	Factor 1	Factor 2
$X_1$	0.687	0.375
$X_2$	0.654	0.467
$X_3$	0.651	0.464
$X_4$	0.708	-0.385
$X_5$	0.688	-0.415
$X_6$	0.710	-0.432



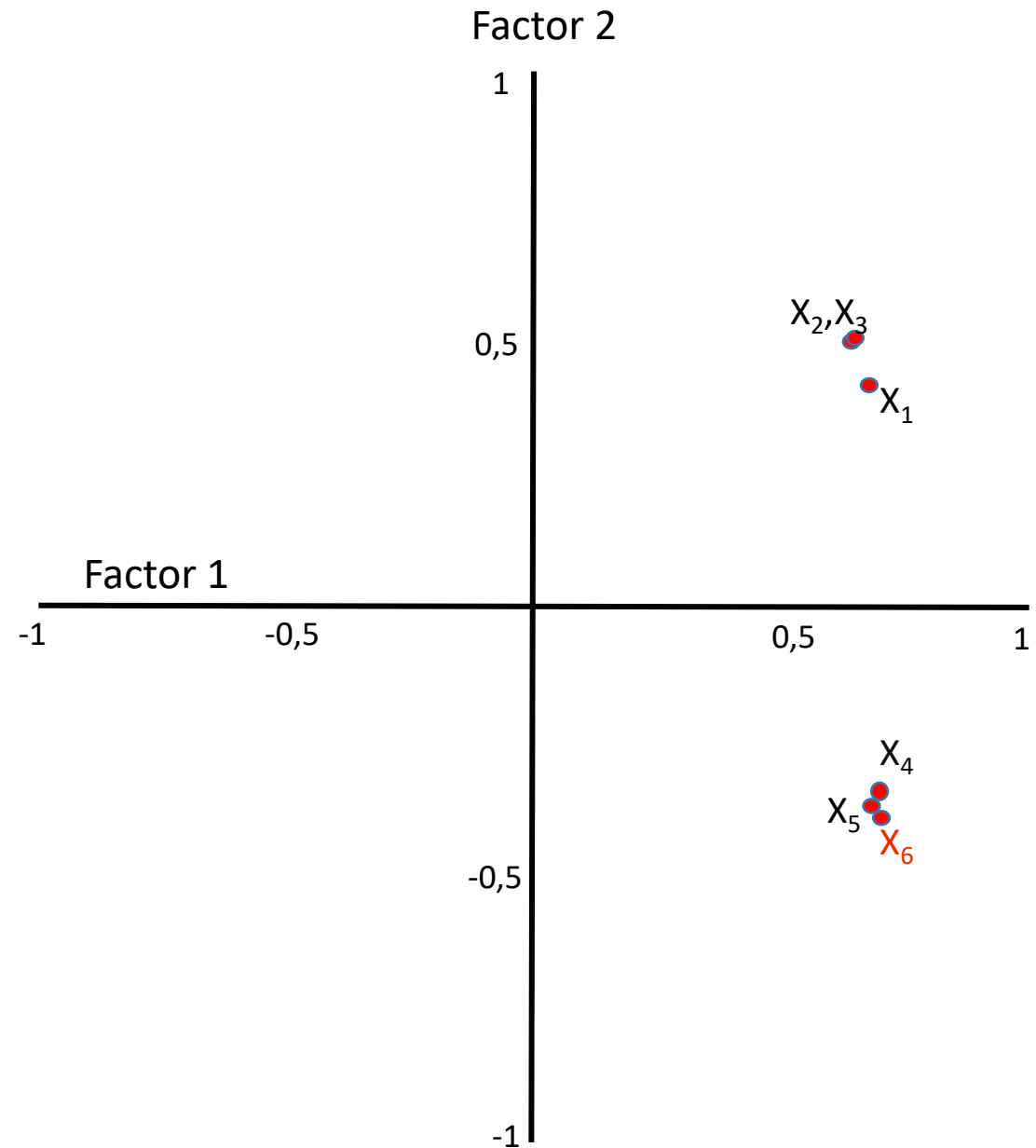
- Factor analysis with 2 factors, based on a dataset with 6 variables ( $X_1, X_2, X_3, X_4, X_5$  &  $X_6$ )
- X axis = loading of every variable on the first factor (factor 1)
- Y axis = loading of every variable on the second factor (factor 2)





- Now: high loadings (in the absolute sense) on both factors
- For example:
- Variable  $X_6$ : loading of  $\sim -.432$  on component 2 and loading of  $.710$  factor 1

-> difficult interpretation



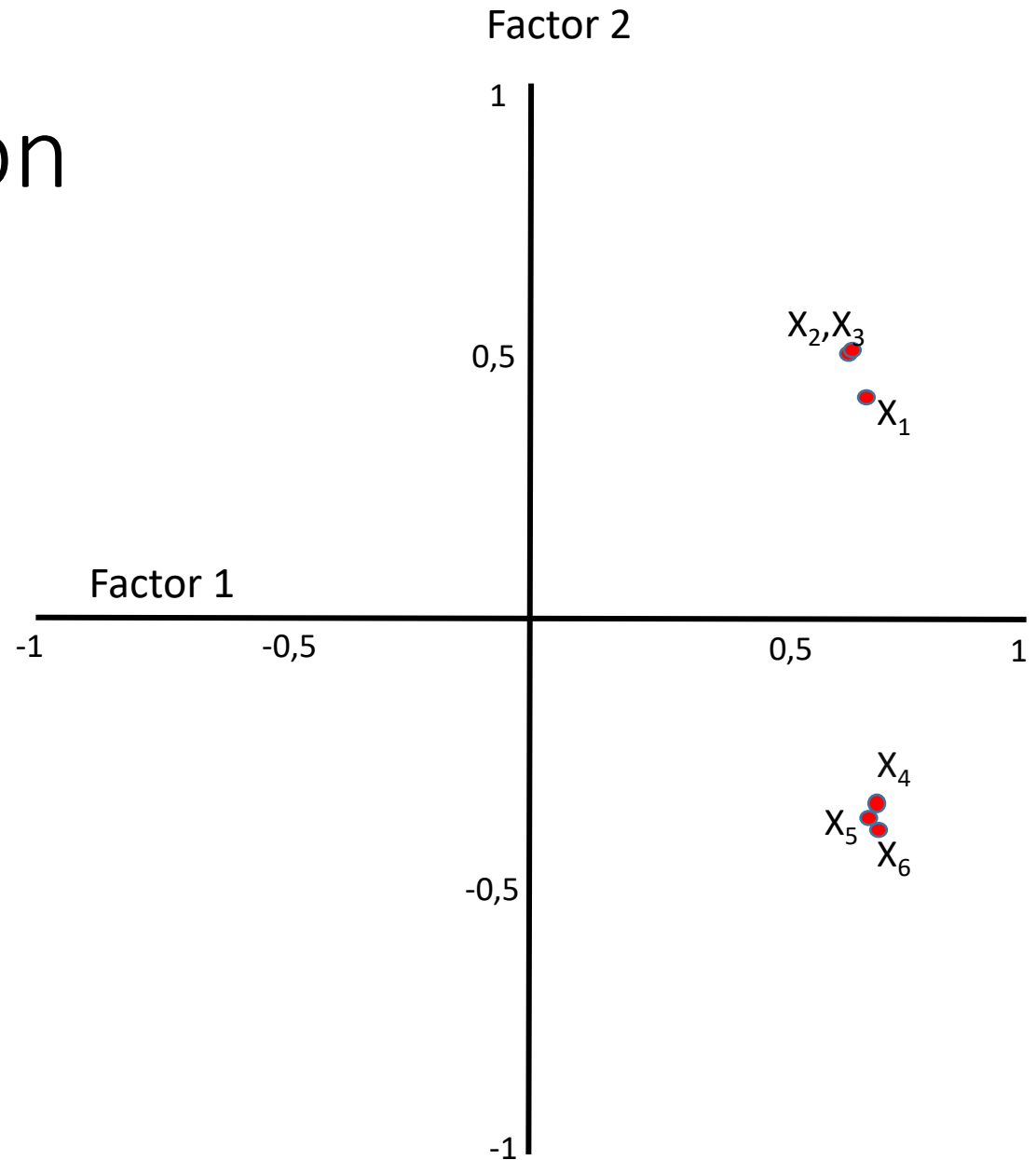
# Rotation: demonstration

## Rotation

*= changing the orientation of the x axis  
and y axis*

This changes the loadings

**The quality of the solution remains the  
same, but the interpretation gets  
easier.**

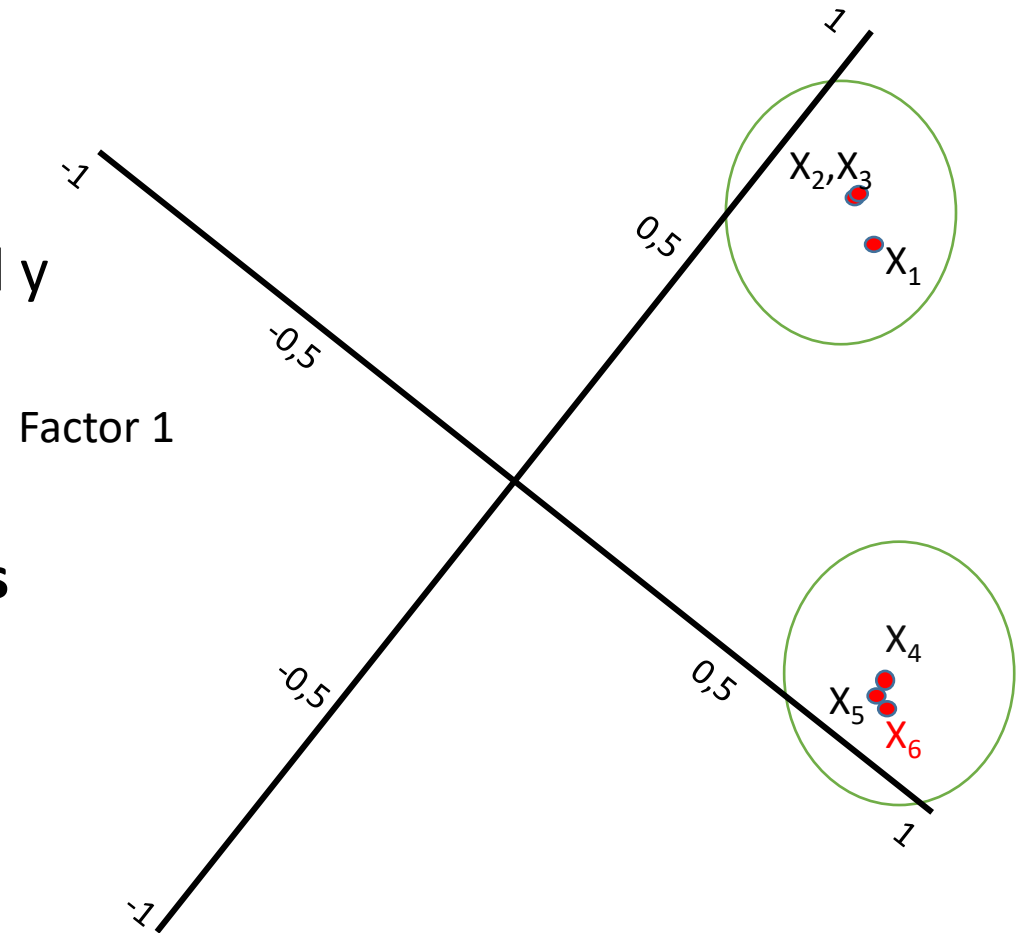


# Rotation:demonstration

## Rotation

= changing the orientation of the x axis and y axis

-> now we can clearly see **2 groups of items**  
e.g., loadings of  $\sim .8$  on factor 1 and  $.1$  on factor 2 for item 6



# Rotation: Summary of what happens

Question: I am confused, what is this rotation about?

Answer: We are interested in another/alternative (easier to interpret) solution.

We want items that follow a **simple structure**, meaning that an item loads strongly only on **one** factor (rule of thumb: loading of  $> |,30|$  with only *one* factor)

Through rotation, the loadings change such that we get closer to a simple structure, but the quality of the solution (the results of the analysis) remain the same (we still explain as much of the variance as we did before)

Pragmatically said:

Rotation = “***Changing***” the loadings such that we can interpret the results easier

Different sorts of rotations:

Orthogonal rotation (“VARIMAX” rotation)

Oblique rotation (“OBLIMIN” rotation)

Two types of rotation:

- **Orthogonal:**

Factors *cannot* correlate with each other:  $r_{F_m F_n} = 0$

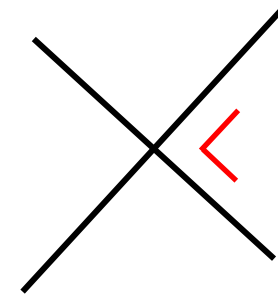
Many different techniques, but most commonly used: **VARIMAX** rotation

Two types of rotation:

- **Orthogonal:**

Factors *cannot* correlate with each other:  $r_{F_m F_n} = 0$

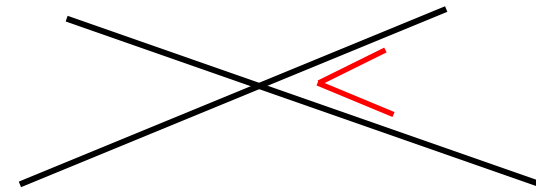
Many different techniques, but most commonly used: **VARIMAX** rotation



- **Oblique:**

Factors *can* correlate with each other:  $r_{F_m F_n} \neq 0$

Many different techniques, but most commonly used: **OBLIMIN** rotation



# 1. VARIMAX



# In SPSS: VARIMAX

Component Matrix <sup>a</sup>

	Component	
	1	2
X1: Anxious	,687	,375
X2: Tense	,654	,467
X3: Restless	,651	,464
X4: Depressed	,708	-,385
X5: Useless	,688	-,415
X6: Unhappy	,710	-,432

Extraction Method: Principal Component Analysis.

a. 2 components extracted.



Rotated Component Matrix <sup>a</sup>

	Component	
	1	2
X1: Anxious		,745
X2: Tense		,789
X3: Restless		,784
X4: Depressed	,779	
X5: Useless	,785	
X6: Unhappy	,813	

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

- Simple structure?
- Easier interpretation than before rotation
- Reflects also what we expected from the correlation matrix (2 groups of items that have high inter-correlations and not very correlated with items of other group)

**Rotated Component Matrix <sup>a</sup>**

	Component	
	1	2
X1: Anxious		,745
X2: Tense		,789
X3: Restless		,784
X4: Depressed	,779	
X5: Useless	,785	
X6: Unhappy	,813	

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

## 2. OBLIMIN

# In SPSS: OBLIMIN

**Pattern Matrix**<sup>a</sup>

	Component	
	1	2
X1: Anxious	,076	,747
X2: Tense	-,030	,816
X3: Restless	-,030	,812
X4: Depressed	,792	,030
X5: Useless	,808	-,010
X6: Unhappy	,837	-,014

Extraction Method: Principal  
Component Analysis.

Rotation Method: Oblimin with Kaiser  
Normalization.

a. Rotation converged in 5 iterations.

- Loadings are now in *pattern* matrix instead of (rotated) component matrix
- Stronger *simple structure* than with VARIMAX (loadings are more extreme / closer to one and zero)

# In SPSS: OBLIMIN

**Component Correlation Matrix**

Component	1	2
1	1,000	,438
2	,438	1,000

Extraction Method: Principal  
Component Analysis.

Rotation Method: Oblimin with  
Kaiser Normalization.

- An assumption of the Oblimin rotation is that there is also a correlation between factors.

So this correlation will be estimated (see table on the left side):

$$r_{F1F2} = 0,438$$

- Which solution to choose?
  - Simple structure => *always rotate*
  - VARIMAX or OBLIMIN? Choose OBLIMIN if
    - Simple structure better attained by OBLIMIN
    - OR, if the correlation between at least one pair of factors is  $\geq |0,30|$
    - Else VARIMAX (easier to work with uncorrelated/unrelated dimensions)
- What kind of rotation would we choose for our example data?

- Ok, so now we know how to interpret the output of a factor analysis in SPSS.
- But, since we are performing a statistical technique, aren't there assumptions we have to check first?!
- -> Yes!

# When is it allowed to perform FA?

- If
  - 1) the assumptions are met
  - 2)  $N$  and  $J$  satisfy some rules ( $N$  = total participants,  $J$  = total items)
  - 3)  $R$  satisfies some rules ( $R$  = correlation matrix of items)



- **1) Assumptions FA**

- Linear relation between the item pairs
- Each item is of an ordinal measurement level with at least 5 answer categories

- 2)  **$N$  and  $J$  ( $N$  = total participants,  $J$  = total items)**

- $N \geq 100$

- $N \geq 5J$

- **3) Correlation matrix  $R$**
- Items have to correlate sufficiently
  - (a) Bartlett's sphericity test must be significant ( $p < 0,05$ )
  - (b) KMO index must be larger than 0,6 ***This is the most important rule!***

# Recap of our example data:

- Measuring depression
- 6 items: *Anxious, tense, restless, depressed, useless & unhappy*
- 300 respondents
- Likert items with 7 answer categories

# Applied to our example data

- 1) Assumptions FA
  - Items have scores ranging from 1 upto and including 7
  - Relations are linear (or, they are not of another form): Check scatter plots!!  
⇒ Correlation is a good measure of association
- 2) Assumptions  $N$  and  $J$ 
  - $N = 300 \geq 100$
  - $N = 300 \geq 5 * J = 30$
- 3)  $R$ 
  - $p < 0,05$
  - $KMO = 0,794$

### Correlation Matrix

		X1: Anxious	X2: Tense	X3: Restless	X4: Depressed	X5: Useless	X6: Unhappy
Correlation	X1: Anxious	1,000	,449	,443	,296	,314	,326
	X2: Tense	<u>,449</u>	1,000	,446	,312	,264	,250
	X3: Restless	<u>,443</u>	<u>,446</u>	1,000	,279	,258	,282
	X4: Depressed	,296	<u>,312</u>	,279	1,000	,467	,516
	X5: Useless	<u>,314</u>	,264	,258	<u>,467</u>	1,000	,497
	X6: Unhappy	<u>,326</u>	,250	,282	<u>,516</u>	<u>,497</u>	1,000

### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		<u>,794</u>
Bartlett's Test of Sphericity	Approx. Chi-Square	430,306
	df	15
	Sig.	<u>,000</u>

# Very important!

- Factor analysis only works well if all assumptions are met & your sample size is big enough. Otherwise, you will get results from SPSS, but they can be biased.
- Hence, check your data for the assumptions
- If the assumptions are not met => **do not apply FA**

- We've heard so many things about reliability and validity...

How, and in what order, are all these steps applied in research practice?

- -> Step by step guide for the complete procedure of exploratory construction of scales



- > Use of all methods in practice
- > Step by step:  
What do we do in research practice?

# Overview of all steps



**Determine validity**

# Overview of all steps

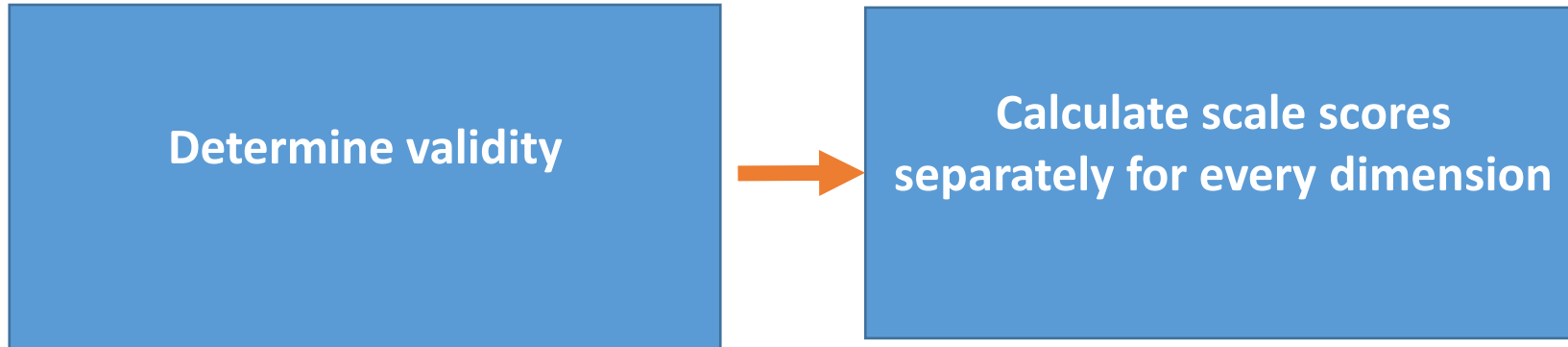
## Determine validity

- 1) How many dimensions?
- 2) Which items belong to what dimension?

-> By using **factor analysis**, we know which items belong to what dimension

The items that measure together one dimension (e.g., a certain aspect of the construct) form together a scale

# Overview of all steps



- 1) How many dimensions?
- 2) Which items belong to what dimension?

-> By using **factor analysis**, we know which items belong to what dimension

The items that measure together one dimension (e.g., a certain aspect of the construct) form together a scale

# Overview of all steps

Determine validity



Calculate scale scores  
separately for every dimension

- 1) How many dimensions?
- 2) Which items belong to what dimension?

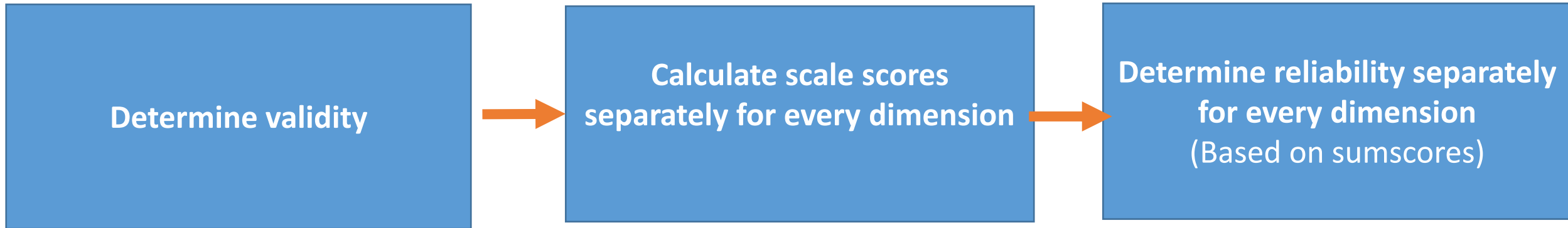
-> By using **factor analysis**, we know which items belong to what dimension

The items that measure together one dimension (e.g., a certain aspect of the construct) form together a scale

Per *dimension* we calculate, for every respondent, a scale score. We do this by using **sumscores**.

The sumscores are based on the answers of a respondent to all items that belong to the dimension (the scale).

# Overview of all steps



- 1) How many dimensions?
- 2) Which items belong to what dimension?

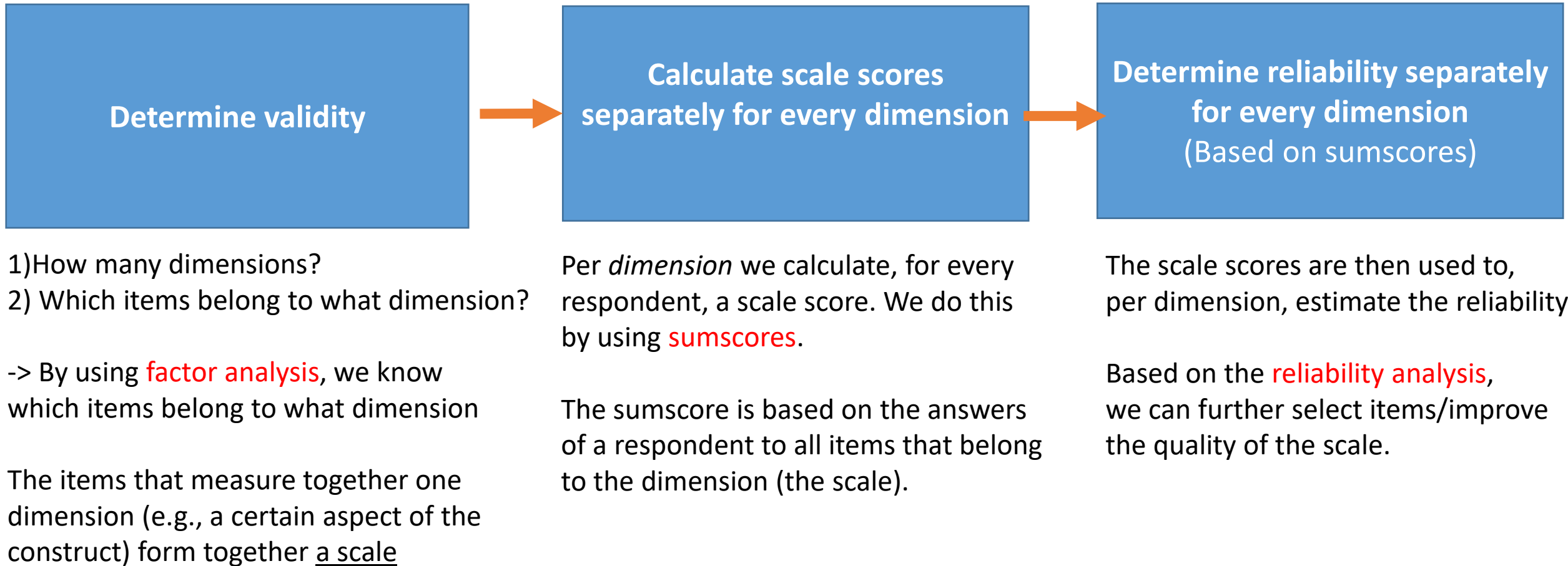
-> By using **factor analysis**, we know which items belong to what dimension

The items that measure together one dimension (e.g., a certain aspect of the construct) form together a scale

Per *dimension* we calculate, for every respondent, a scale score. We do this by using **sumscores**.

The sumscores are based on the answers of a respondent to all items that belong to the dimension (the scale).

# Overview of all steps



# Focus of today: Factor analysis



Determine validity

- 1) How many dimensions?
- 2) Which items belong to what dimension?

-> By using **factor analysis**, we know which items belong to what dimension

The items that measure together one dimension (e.g., a certain aspect of the construct) form together a scale



# Factor analysis: all steps

- 1) Are the assumptions met?
- 2) Rotation: VARIMAX or OBLIMIN?
- 3) Interpretation of the components/factors
  - E.g., component/factor 1: Which dimension is that? –  
-> Look at the content of the items & choose a label based on that

# Exploratory FA: Complete procedure

- For a complete FA, we need to perform 5 steps:
  - (i) **Check if the association between items is linear**

# Exploratory FA: Complete procedure

- Five steps / analyses:
  - (i) **Check if the association between items is linear**
  - (ii) **First FA: does it make sense to perform FA + how many factors/components?**
    - KMO & Bartlett's sphericity test – assumptions met?
    - How many factors do we choose? (How many dimensions?)

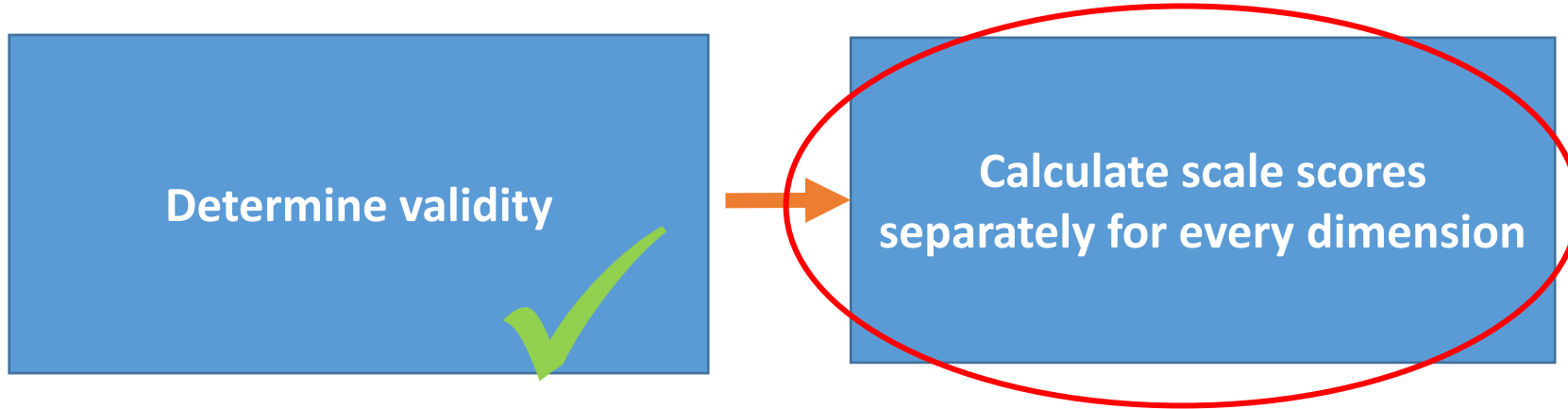
# Exploratory FA: Complete procedure

- Five steps / analyses:
  - (i) **Check if the association between items is linear**
  - (ii) **First FA: does it make sense to perform FA + how many factors/components?**
    - KMO & Bartlett's sphericity test – assumptions met?
    - How many factors do we choose? (How many dimensions?)
  - (iii) **Second FA: VARIMAX**
  - (iv) **Third FA: OBLIMIN**

# Exploratory FA: Complete procedure

- Five steps / analyses:
  - (i) **Check if the association between items is linear**
  - (ii) **First FA: does it make sense to perform FA + how many factors/components?**
    - KMO & Bartlett's sphericity test – assumptions met?
    - How many factors do we choose? (How many dimensions?)
  - (iii) **Second FA: VARIMAX**
  - (iv) **Third FA: OBLIMIN**
  - (v) **Choose a solution between (iii) & (iv) and interpret it**
    - Which solution is easier to interpret? + is there a correlation between factors?
    - After choosing: Interpretation of the dimensions -> look at content of the items

# Once we have determined the internal structure, we have to calculate scale scores



- 1) How many dimensions?
- 2) Which items belong to what dimension?

-> By using **factor analysis**, we know which items belong to what dimension

The items that measure together one dimension (e.g., a certain aspect of the construct) form together a scale

Per *dimension* we calculate, for every respondent, a scale score. We do this by using **sumscores**.

The sumscores are based on the answers of a respondent to all items that belong to the dimension (the scale).

# Calculation of a sumscore in practice

- Sum score => sum up only those items that form one scale (= that belong to the same dimension)
- For example thus, for a questionnaire consisting of 10 items :  
Dimension 1 -> Item 1,2,4,5,7 = SCALE 1 -> sumscore 1  
Dimension 2 -> Item 3,6,8,9,10 = SCALE 2 -> sumscore 2

From where can we get this information??!!

-> last step of the FA-steps that we just dicussed:

the output gives us information on which items form together one scale  
(e.g. – decide for every item to which dimension/scale this item belongs - based on the loadings)

- Imagine that we have applied the 5 steps plan and conducted a PCA.

- To the right: component loadings

⇒ Dimension 1: X4, X5, X6 = SCALE 1

⇒ Dimension 2: X1, X2, X2 = SCALE 2

⇒ So:

$X4 + X5 + X6 = \text{SUMSCORE } 1,$

$X1 + X2 + X2 = \text{SUMSCORE } 2$

For every respondent, we then calculate 2 sumscores.

**Rotated Component Matrix <sup>a</sup>**

	Component	
	1	2
X1: Anxious		,745
X2: Tense		,789
X3: Restless		,784
X4: Depressed	,779	
X5: Useless	,785	
X6: Unhappy	,813	

Extraction Method: Principal

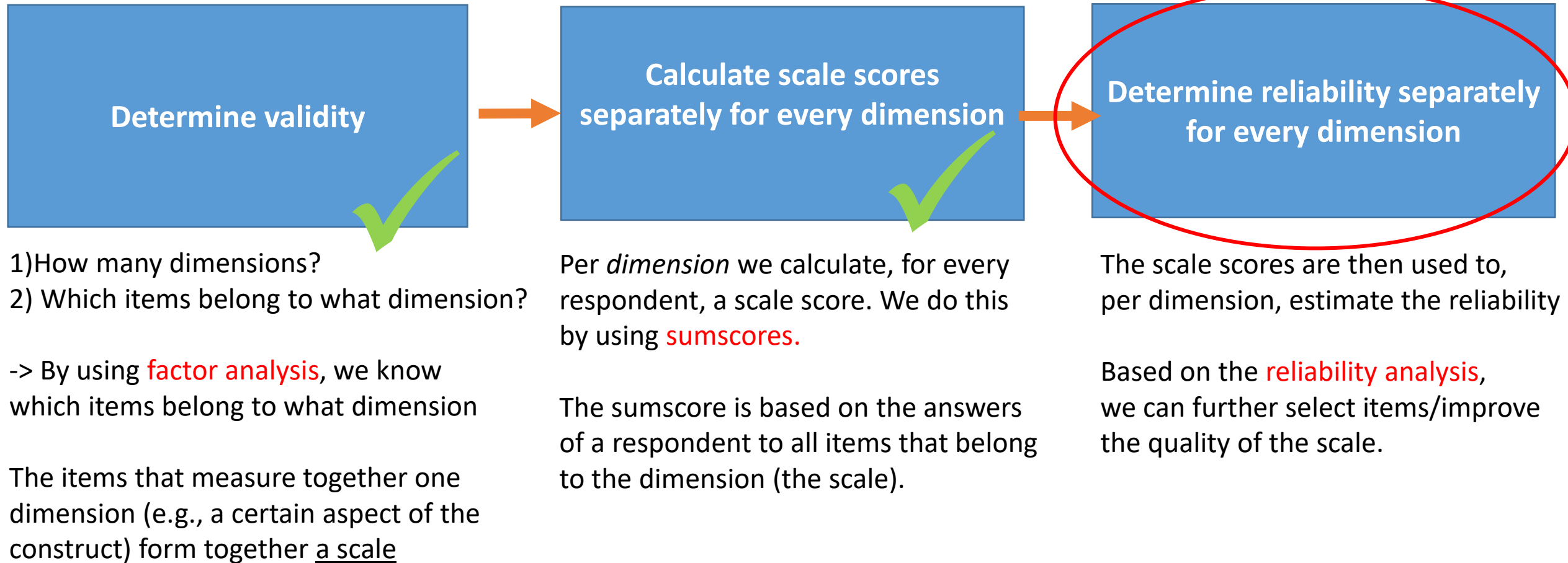
Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.



# Based on the scale scores (sumscores), we determine the reliability of every separate scale



# Based on the sumscores: determine reliability

- So we have to calculate 2 sumscores for every respondent.
- Then, we perform for every scale, based on the sumscores that belong to that particular scale, a reliability analysis
- *Thus, per scale* we run one analysis and one Cronbach's alpha
- This is what you have done/learned yesterday

For the example data:  
X1,X2 en X3 form together one scale and X4, X5  
and X6

**Reliability Statistics**

Cronbach's Alpha	N of Items
,706	3

**Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
X1: Anxious	7,8967	3,003	,524	,615
X2: Tense	8,0300	3,073	,527	,611
X3: Restless	8,1400	3,365	,522	,620

**Reliability Statistics**

Cronbach's Alpha	N of Items
,745	3

**Item-Total Statistics**

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
X4: Depressed	7,9267	3,239	,569	,663
X5: Useless	7,8867	3,499	,553	,681
X6: Unhappy	7,8933	3,206	,592	,635

# All steps: see next slide

## Determine validity

- 1) How many dimensions?
- 2) Which items belong to what dimension?

-> By using **factor analysis**, we know which items belong to what dimension

The items that measure together one dimension (e.g., a certain aspect of the construct) form together a scale

## Calculate scale scores separately for every dimension

Per *dimension* we calculate, for every respondent, a scale score. We do this by using **sumscores**.

The sumscore is based on the answers of a respondent to all items that belong to the dimension (the scale).

## Determine reliability separately for every dimension

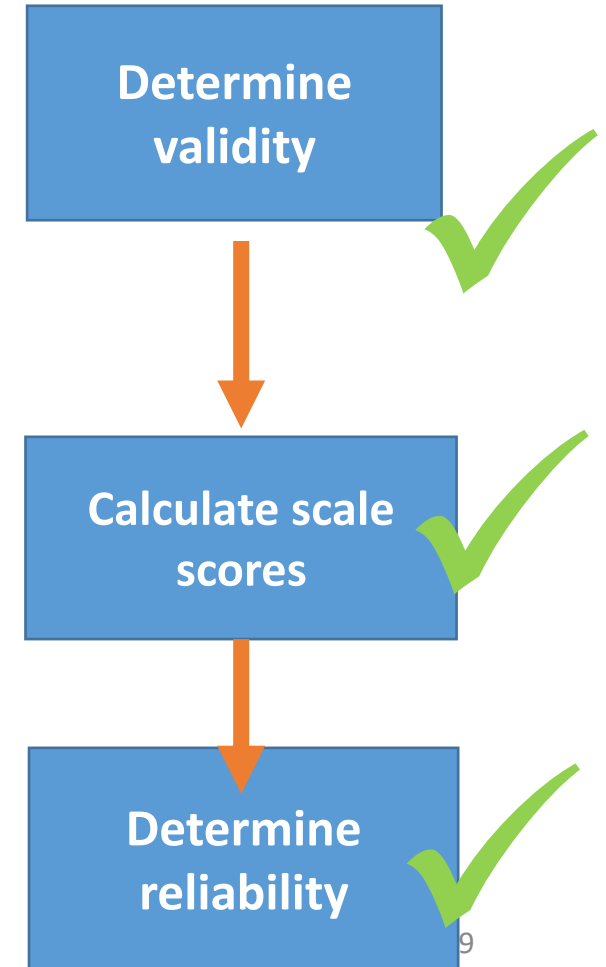
The scale scores are then used to, per dimension, estimate the reliability

Based on the **reliability analysis**, we can further select items/improve the quality of the scale.

# And these are all steps:

## Exploratory **construction of scales**

1. Check if the association between items is linear
2. First FA: does it make sense to perform PFA + how many factors/components do we have?
3. Second FA: VARIMAX
4. Third FA: OBLIMIN
5. Choose a solution between 3 & 4 and interpret it
6. Calculate sum scores for each scale / dimension
7. Evaluate **reliability** of every scale, the item contributions to the reliability (per scale), and report Cronbach's alpha per scale



# Advanced topics

- Some more advanced analyses are not possible in SPSS
  - **Test internal structure that is known beforehand**  
(*Confirmatory factor analysis*) -> see example script on github
  - **Test for measurement invariance**  
Compare multiple groups (e.g., companies/countries)  
(*Multiple group analysis*)  
*Not possible by conducting separate factor analyses!!!*