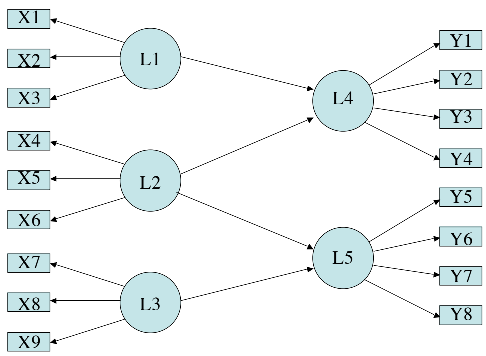
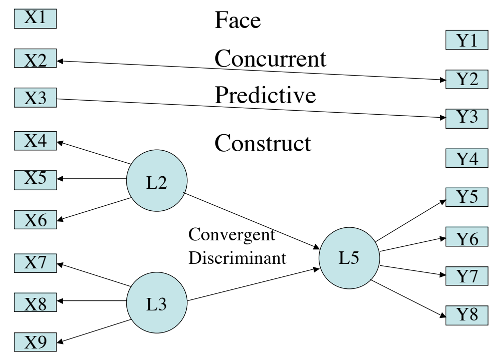
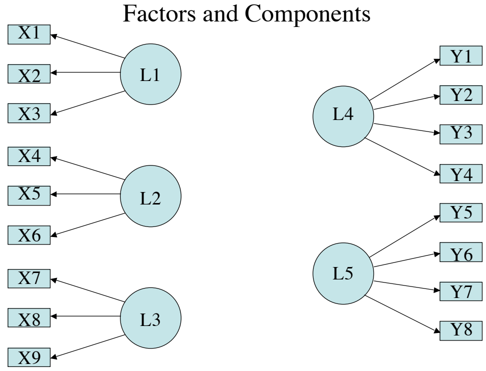
Psychometric Theory: A conceptual Syllabus




Types of Validity: What are we measuring




Techniques of Data Reduction:
Factors and Components




# Methods of Homogeneous Keying

Factor analysis, principal components analysis, and cluster analysis

# Methods of Homogeneous Keying

- Factor Analysis
- Principal Components Analysis
- Cluster Analysis

# Factor Analysis
Consider the following r matrix

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| $X_1$ | 1.00 | | | | |
| $X_2$ | .72 | 1.00 | | | |
| $X_3$ | .63 | .56 | 1.00 | | |
| $X_4$ | .54 | .48 | .42 | 1.00 | |
| $X_5$ | .45 | .40 | .35 | .30 | 1.00 |

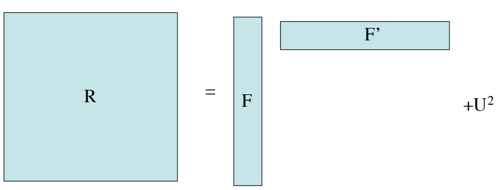
## Factors as causes; variables as indicators of factors

- Can we recreate the correlation matrix R (of rank n) with a matrix F of rank 1 + a diagonal matrix of uniqueness $U^2$
- $R \approx FF' + U^2$
- Residual Matrix R* = R - (FF'+ U2)
- Try to minimize the residual

## 1 Factor solution $R \approx FF' + U^2$

| | Correlation with (loading on) Factor 1 |
|---|---|
| $X_1$ | .9 |
| $X_2$ | .8 |
| $X_3$ | .7 |
| $X_4$ | .6 |
| $X_5$ | .5 |

## Factor analysis




## Factor Analysis more than 1 factor

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| $X_1$ | 1.00 | | | | |
| $X_2$ | .72 | 1.00 | | | |
| $X_3$ | .00 | .00 | 1.00 | | |
| $X_4$ | .00 | .00 | .42 | 1.00 | |
| $X_5$ | .00 | .00 | .35 | .30 | 1.00 |

## Factor Analysis: the model

- $R \approx FF' + U^2$
- Residual Matrix R* = R - (FF'+ U2)
- Try to minimize the residual
- Variables are linear composites of unknown (latent) factors.
- Covariance structures of observables in terms of covariance of unobservables

## Structure of mood - how not to display data

| | AFRAID | AT_EASE | CALM | ENERGETI | HAPPY | LIVELY | SLEEPY | TENSE | TIRED |
|---|---|---|---|---|---|---|---|---|---|
| AFRAID | 1.000 | | | | | | | | |
| AT_EASE | -0.209 | 1.000 | | | | | | | |
| CALM | -0.157 | 0.586 | 1.000 | | | | | | |
| ENERGETI | 0.019 | 0.230 | 0.056 | 1.000 | | | | | |
| HAPPY | -0.070 | 0.452 | 0.294 | 0.595 | 1.000 | | | | |
| LIVELY | 0.018 | 0.255 | 0.073 | 0.778 | 0.609 | 1.000 | | | |
| SLEEPY | 0.087 | -0.112 | 0.031 | -0.457 | -0.264 | -0.405 | 1.000 | | |
| TENSE | 0.397 | -0.337 | -0.332 | 0.088 | -0.103 | 0.084 | 0.044 | 1.000 | |
| TIRED | 0.082 | -0.141 | 0.012 | -0.484 | -0.297 | -0.439 | 0.808 | 0.044 | 1.000 |
| UNHAPPY | 0.350 | -0.283 | -0.187 | -0.185 | -0.314 | -0.187 | 0.202 | 0.360 | 0.235 |

## Structure of mood: “Alabama need come first”

| | AFRAID | AT_EASE | CALM | ENERGETI | HAPPY | LIVELY | SLEEPY | TENSE | TIRED |
|---|---|---|---|---|---|---|---|---|---|
| AFRAID | 1.0 | | | | | | | | |
| AT_EASE | -0.2 | 1.0 | | | | | | | |
| CALM | -0.2 | 0.6 | 1.0 | | | | | | |
| ENERGETI | 0.0 | 0.2 | 0.1 | 1.0 | | | | | |
| HAPPY | -0.1 | 0.5 | 0.3 | 0.6 | 1.0 | | | | |
| LIVELY | 0.0 | 0.3 | 0.1 | 0.8 | 0.6 | 1.0 | | | |
| SLEEPY | 0.1 | -0.1 | 0.0 | -0.5 | -0.3 | -0.4 | 1.0 | | |
| TENSE | 0.4 | -0.3 | -0.3 | 0.1 | -0.1 | 0.1 | 0.0 | 1.0 | |
| TIRED | 0.1 | -0.1 | 0.0 | -0.5 | -0.3 | -0.4 | 0.8 | 0.0 | 1.0 |
| UNHAPPY | 0.3 | -0.3 | -0.2 | -0.2 | -0.3 | -0.2 | 0.2 | 0.4 | 0.2 |

## Structure of mood data

| | ENERGETI | LIVELY | TIRED | SLEEPY | AFRAID | TENSE | AT_EASE | CALM | HAPPY |
|---|---|---|---|---|---|---|---|---|---|
| ENERGETI | 1 | | | | | | | | |
| LIVELY | 0.8 | 1 | | | | | | | |
| TIRED | -0.5 | -0.4 | 1 | | | | | | |
| SLEEPY | -0.5 | -0.4 | 0.8 | 1 | | | | | |
| AFRAID | 0 | 0 | 0.1 | 0.1 | 1 | | | | |
| TENSE | 0.1 | 0.1 | 0 | 0 | 0.4 | 1 | | | |
| AT_EASE | 0.2 | 0.3 | -0.1 | -0.1 | -0.2 | -0.3 | 1 | | |
| CALM | 0.1 | 0.1 | 0 | 0 | -0.2 | -0.3 | 0.6 | 1 | |
| HAPPY | 0.6 | 0.6 | -0.3 | -0.3 | -0.1 | -0.1 | 0.5 | 0.3 | 1 |
| UNHAPPY | -0.2 | -0.2 | 0.2 | 0.2 | 0.3 | 0.4 | -0.3 | -0.2 | -0.3 |

NUMBER OF OBSERVATIONS: 3748

## Correlation of mood data possible structure

| | ENERGETI | LIVELY | TIRED | SLEEPY | AFRAID | TENSE | AT_EASE | CALM | HAPPY |
|---|---|---|---|---|---|---|---|---|---|
| ENERGETI | 1.0 | | | | | | | | |
| LIVELY | 0.8 | 1.0 | | | | | | | |
| TIRED | -0.5 | -0.4 | 1.0 | | | | | | |
| SLEEPY | -0.5 | -0.4 | 0.8 | 1.0 | | | | | |
| AFRAID | 0.0 | 0.0 | 0.1 | 0.1 | 1.0 | | | | |
| TENSE | 0.1 | 0.1 | 0.0 | 0.0 | 0.4 | 1.0 | | | |
| AT_EASE | 0.2 | 0.3 | -0.1 | -0.1 | -0.2 | -0.3 | 1.0 | | |
| CALM | 0.1 | 0.1 | 0.0 | 0.0 | -0.2 | -0.3 | 0.6 | 1.0 | |
| HAPPY | 0.6 | 0.6 | -0.3 | -0.3 | -0.1 | -0.1 | 0.5 | 0.3 | 1.0 |
| UNHAPPY | -0.2 | -0.2 | 0.2 | 0.2 | 0.3 | 0.4 | -0.3 | -0.2 | -0.3 |

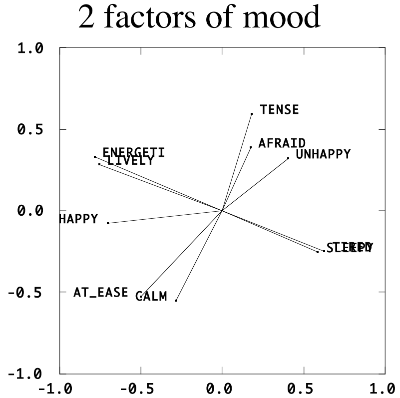
NUMBER OF OBSERVATIONS: 3748

## Factor analysis 2 factor solution

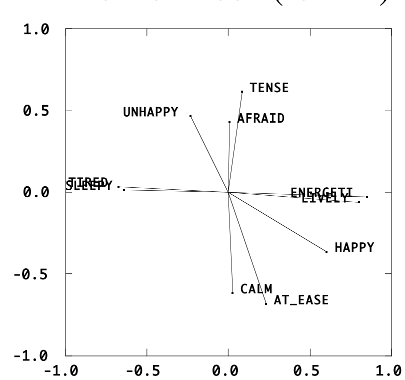
FACTOR PATTERN

| | 1 | 2 | h2 |
|---|---|---|---|
| ENERGETI | -0.8 | 0.3 | 0.73 |
| LIVELY | -0.8 | 0.3 | 0.73 |
| HAPPY | -0.7 | -0.1 | 0.50 |
| TIRED | 0.6 | -0.3 | 0.45 |
| SLEEPY | 0.6 | -0.3 | 0.45 |
| TENSE | 0.2 | 0.6 | 0.40 |
| CALM | -0.3 | -0.6 | 0.45 |
| AT_EASE | -0.5 | -0.5 | 0.50 |
| AFRAID | 0.2 | 0.4 | 0.20 |
| UNHAPPY | 0.4 | 0.3 | 0.25 |
| VARIANCE EXP | 3.0 | 1.50 | |

## 2 factors of mood



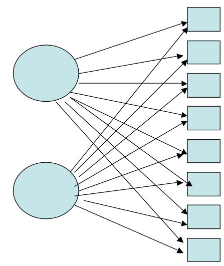

## 2 factors of mood (rotated)



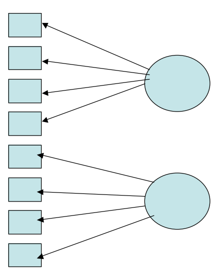
## Rotation as orthogonal transformation

| | F1 | F2 | F1' | F2' |
|---|---|---|---|---|
| ENERGETI | -0.8 | 0.3 | 0.8 | 0.0 |
| LIVELY | -0.8 | 0.3 | 0.8 | -0.1 |
| HAPPY | -0.7 | -0.1 | 0.6 | -0.4 |
| TIRED | 0.6 | -0.3 | -0.7 | 0.0 |
| SLEEPY | 0.6 | -0.3 | -0.6 | 0.0 |
| TENSE | 0.2 | 0.6 | 0.1 | 0.6 |
| CALM | -0.3 | -0.6 | 0.0 | -0.6 |
| AT_EASE | -0.5 | -0.5 | 0.2 | -0.7 |
| AFRAID | 0.2 | 0.4 | 0.0 | 0.4 |
| UNHAPPY | 0.4 | 0.3 | -0.2 | 0.5 |
| eigen values | 3 | 1.5 | 2.7 | 1.8 |

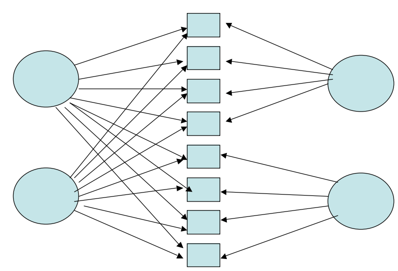
## Rotation to simple structure Original



## Rotation to simple structure Simple Structure



## Rotation to simple structure Original Rotated



## Principal components

- R≈ CC’
- Residual Matrix R* = R - (CC’)
- Try to minimize the residual
- Components are linear composites of known variables.
- Covariance structures of observables in terms of covariance of observables
- Components account for observed variance

## 2 Principal Components of mood

| | C1 | C2 |
|---|---|---|
| ENERGETI | -0.8 | -0.4 |
| LIVELY | -0.8 | -0.3 |
| HAPPY | -0.8 | 0 |
| TIRED | 0.7 | 0.3 |
| SLEEPY | 0.7 | 0.3 |
| AT_EASE | -0.6 | 0.5 |
| TENSE | 0.2 | -0.7 |
| CALM | -0.3 | 0.6 |
| AFRAID | 0.2 | -0.5 |
| UNHAPPY | 0.5 | -0.4 |
| eigen values | 3.4 | 2.1 |

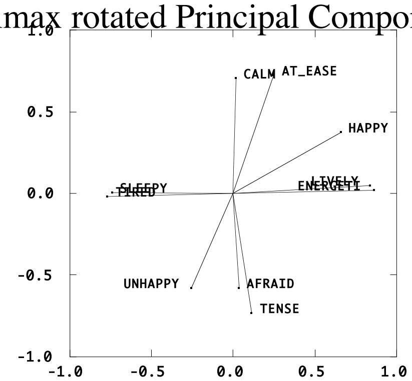
## Unrotated and rotated PCs

| | C1 | C2 | C1' | C2' |
|---|---|---|---|---|
| ENERGETI | -0.8 | -0.4 | 0.9 | 0 |
| LIVELY | -0.8 | -0.3 | 0.8 | 0 |
| HAPPY | -0.8 | 0 | 0.7 | 0.4 |
| TIRED | 0.7 | 0.3 | -0.8 | 0 |
| SLEEPY | 0.7 | 0.3 | -0.7 | 0 |
| AT_EASE | -0.6 | 0.5 | 0.2 | 0.7 |
| TENSE | 0.2 | -0.7 | 0.1 | -0.7 |
| CALM | -0.3 | 0.6 | 0 | 0.7 |
| AFRAID | 0.2 | -0.5 | 0 | -0.6 |
| UNHAPPY | 0.5 | -0.4 | -0.3 | -0.6 |
| eigen values | 3.4 | 2.1 | 3.2 | 2.4 |

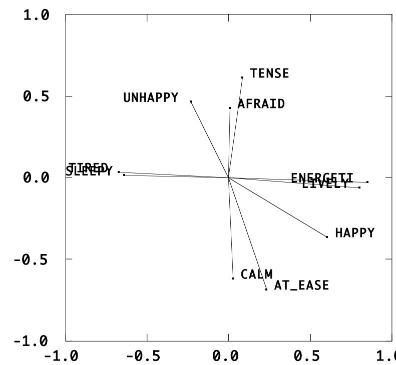
## Varimax factors and components

| | F1' | F2' | C1' | C2' |
|---|---|---|---|---|
| ENERGETI | 0.8 | 0.0 | 0.9 | 0.0 |
| LIVELY | 0.8 | -0.1 | 0.8 | 0.0 |
| HAPPY | 0.6 | -0.4 | 0.7 | 0.4 |
| TIRED | -0.7 | 0.0 | -0.8 | 0.0 |
| SLEEPY | -0.6 | 0.0 | -0.7 | 0.0 |
| AT_EASE | 0.2 | -0.7 | 0.2 | 0.7 |
| TENSE | 0.1 | 0.6 | 0.1 | -0.7 |
| CALM | 0.0 | -0.6 | 0.0 | 0.7 |
| AFRAID | 0.0 | 0.4 | 0.0 | -0.6 |
| UNHAPPY | -0.2 | 0.5 | -0.3 | -0.6 |
| eigen values | 2.7 | 1.8 | 3.2 | 2.4 |

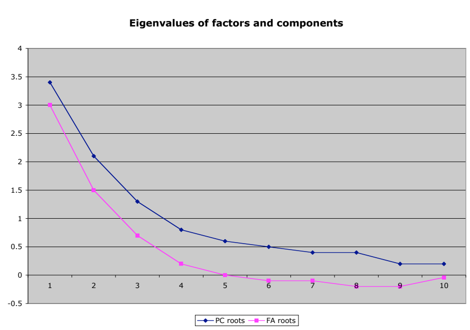
## Varimax rotated Principal Components



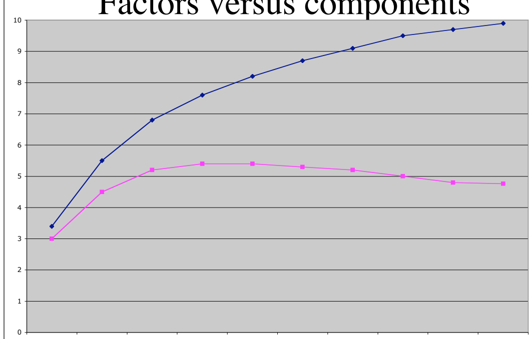

## Varimax rotated factors of mood



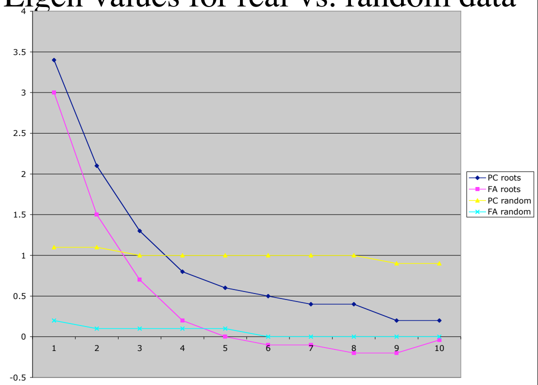

## Eigenvalues and the scree test




## Cumulative variance explained Factors versus components



## Eigen values for real vs. random data




## Determining the number of factors/components

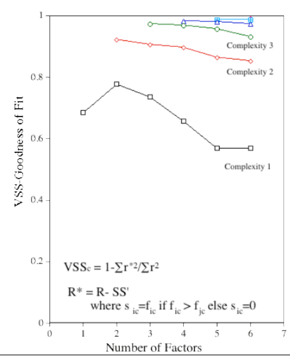
- Statistically: extract factors until residual matrix does not differ from 0
  - Sensitive to sample size (large N -> more factors)
- Theoretically: extract factors that make sense
  - Different theorists differ in their cognitive complexity
- Pragmatically: scree test
- Pragmatically: extract factors with eigen values greater than a random factor matrix
- Pragmatically: extract factors using Very Simple Structure Test or the minimal partial correlation
- Pragmatically: do not use eigen value>1 rule!

## Very Simple Structure

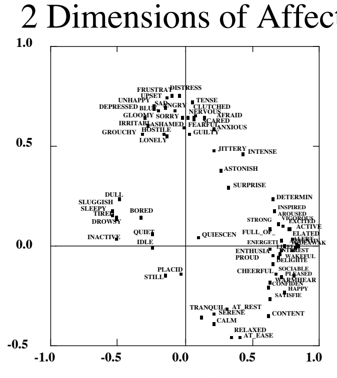
(Revelle and Rocklin)

- Consider the factors as interpreted by user
  - Small loadings are thought to be zero
- How well does this very simple interpretation of the actual structure reproduce the correlation matrix
- $R^*_{vss} = R - F_{vss} * F'_{vss}$
- Consider structures of various levels of simplicity (complexity) (1, 2, … loadings/item)
- Solution peaks at optimal number of factors

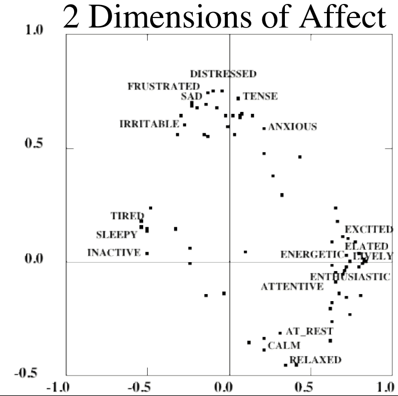
## Very Simple Structure => 2 Factors




## 2 Dimensions of Affect



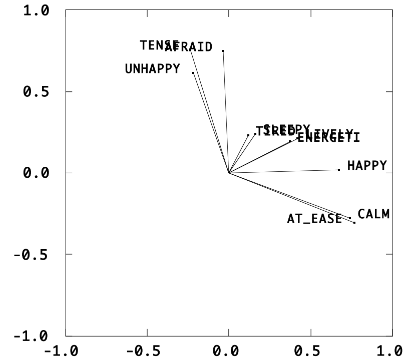

## 2 Dimensions of Affect



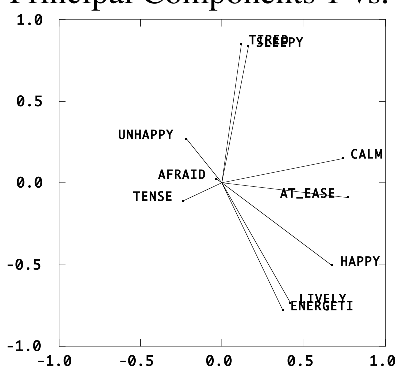
# Representative MSQ items (arranged by angular location)

| Item | EA-PA | TA-NA | Angle |
|---|---|---|---|
| **energetic** | **0.8** | 0.0 | 1 |
| elated | **0.7** | 0.0 | 2 |
| excited | **0.8** | 0.1 | 6 |
| anxious | 0.2 | **0.6** | 70 |
| **tense** | 0.1 | **0.7** | 85 |
| distressed | 0.0 | **0.8** | 93 |
| **frustrated** | -0.1 | **0.8** | 98 |
| sad | -0.1 | **0.7** | 101 |
| irritable | -0.3 | **0.6** | 114 |
| **sleepy** | **-0.5** | 0.1 | 164 |
| **tired** | **-0.5** | 0.2 | 164 |
| inactive | **-0.5** | 0.0 | 177 |
| **calm** | 0.2 | **-0.4** | 298 |
| **relaxed** | 0.4 | **-0.5** | 307 |
| at ease | 0.4 | **-0.5** | 312 |
| attentive | **0.7** | 0.0 | 357 |
| enthusiastic | **0.8** | 0.0 | 358 |
| **lively** | **0.9** | 0.0 | 360 |

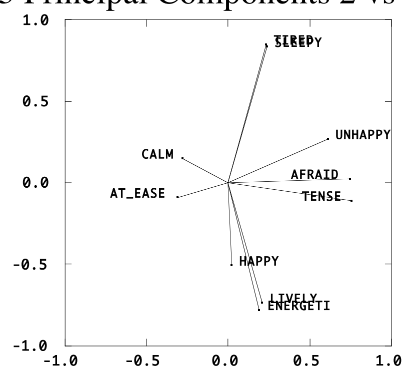
# 3 Principal Components (1 vs. 2)




# 3 Principal Components 1 vs. 3




# 3 Principal Components 2 vs 3




# FA and PCA vocabulary

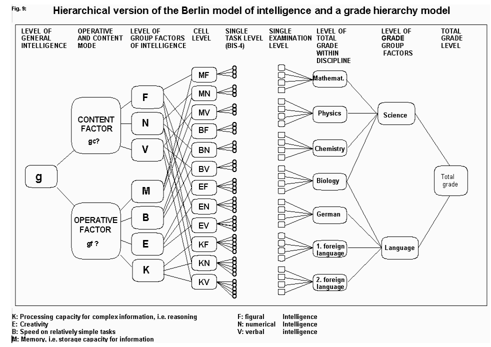
- Eigen values = ∑(loading$^2$) across variables = amount of variance accounted for by factor
- Communality = ∑(loading$^2$) across factors = amount of variance accounted for in a variable by all the factors
- Rotations versus Transformations
  - Rotations are orthogonal transformations
  - Oblique Transformations

# Rotations and transformations

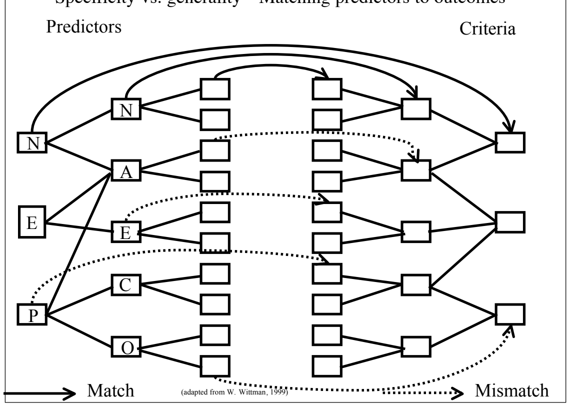
- Simple structure as a criterion for rotation
- Simple structure in the eye of the beholder
- Simple factors (few high, many 0 loadings)
- Simple variables (few high, many 0 loadings)
- VARIMAX, Quartimax, Quartimin
- Procrustes

# Rotations and transformations

- Orthogonal rotations
  - Factors are orthongonal, rotated to reduce (or maximize) particular definition of simple structure
- Oblique transformations and higher order factors
  - Allows factors to be correlated (and thus have higher order factors)




Specificity vs. generality --Matching predictors to outcomes




# Exploratory versus Confirmatory

- Exploratory:
  - How many factors and best rotation
  - Extraction
    - How many factors?
    - Algorithm for extraction
  - Rotation to simple structure -- what is best SS?
- Confirmatory: does a particular model fit?
  - Apply statistical test of fit
  - But larger N => less fit
  - Is the model the original one or has it been modified?

# Components versus Factors

- Components are linear sums of variables and are thus defined at the data level
- Factors represent the covariances of variables and are only estimated (variables are linear sums of factors)
  - Model is undefined at data level, but is defined at structural level
  - Factor indeterminancy problem

# Threats to interpreting correlation and the benefits of structure

- Correlations can be attenuated due to differences in skew (and error)
- Bi polar versus unipolar scales of affect
- How happy do you feel?
  - Not at all   A little   Somewhat   Very
- How sad do you feel?
  - Not at all   A little   Somewhat   Very
- How do you feel?

  very sad   sad   happy   very happy

## Simulated Example of unipolar scales and the problem of skew

- Consider X and Y as random normal variables
- Let X+ = X if X>0, 0 elsewhere
- Let X- = X if X<0, 0 elsewhere
- Reversed (X+) = - X+
- Similarly for Y
- Examine the correlational structure
- Note that although X and -X correlate -1.0, X+ and X- correlate only -.43 and that X+ correlates with X+Y+ .66

## Determining Structure: zeros and skew

| | X+ | X- | Y+ | Y- | X+Y+ | X-Y- | X+Y- | X-Y+ |
|---|---|---|---|---|---|---|---|---|
| X+ | 1.00 | | | | | | | |
| X- | -0.47 | 1.00 | | | | | | |
| Y+ | 0.03 | -0.01 | 1.00 | | | | | |
| Y- | 0.00 | -0.03 | -0.46 | 1.00 | | | | |
| X+Y+ | 0.65 | -0.39 | 0.66 | -0.39 | 1.00 | | | |
| X-Y- | -0.40 | 0.63 | -0.40 | 0.63 | -0.46 | 1.00 | | |
| X+Y- | 0.63 | -0.40 | -0.39 | 0.66 | 0.00 | 0.02 | 1.00 | |
| X-Y+ | -0.39 | 0.64 | 0.63 | -0.40 | 0.00 | 0.00 | -0.47 | 1.00 |

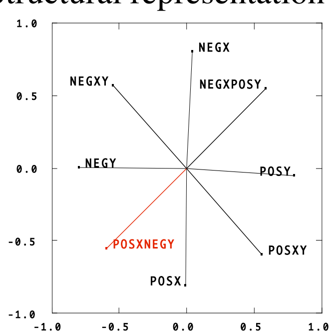
## Skew and zeros: determining structure

| | Y+ | X+Y+ | X+ | X+Y- | r(Y-) | r(X-Y-) | r(X-) | X-Y+ |
|---|---|---|---|---|---|---|---|---|
| Y | 1.00 | | | | | | | |
| X+Y+ | 0.66 | 1.00 | | | | | | |
| X+ | 0.03 | 0.65 | 1.00 | | | | | |
| X+Y- | -0.39 | 0.00 | 0.63 | 1.00 | | | | |
| r(Y-) | 0.46 | 0.39 | 0.00 | -0.66 | 1.00 | | | |
| r(X-Y-) | 0.40 | 0.46 | 0.40 | -0.02 | 0.63 | 1.00 | | |
| r(X-) | 0.01 | 0.39 | 0.47 | 0.40 | -0.03 | 0.63 | 1.00 | |
| X-Y+ | 0.63 | 0.00 | -0.39 | -0.47 | 0.40 | 0.00 | -0.64 | 1.00 |

## Factor analysis shows structure

| | 1 | 2 | |
|---|---|---|---|
| POSX | -0.01 | -0.81 | 271 |
| NEGX | 0.04 | 0.80 | 87 |
| POSY | 0.80 | -0.05 | 356 |
| NEGY | -0.80 | 0.00 | 180 |
| POSXY | 0.56 | -0.59 | 314 |
| NEGXY | -0.55 | 0.57 | 136 |
| POSXN | -0.59 | -0.55 | 227 |
| NEGXP | 0.58 | 0.55 | 43 |

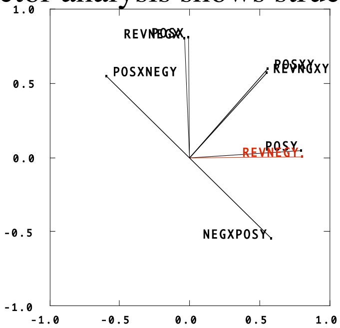
## Structural representation




## Factor analysis shows structure

| | 1 | 2 | angle |
|---|---|---|---|
| POSX | -0.01 | 0.81 | 89 |
| POSY | 0.80 | 0.05 | 4 |
| POSXY | 0.56 | 0.59 | 46 |
| POSXN | -0.59 | 0.55 | 137 |
| NEGXP | 0.58 | -0.55 | 313 |
| REVNE | -0.04 | 0.80 | 87 |
| REVNE | 0.80 | 0.00 | 90 |
| REVNG | 0.55 | 0.57 | 46 |

## Factor analysis shows structure




## Hyperplanes and Zeros: Defining variables by what they are not

- Tendency to interpret variables in terms of their patterns of high correlations
- But large correlations may be attenuated by skew or error
- Correlation of .7 is an angle of 45 degrees => lots of room for misinterpretation!
- Zero correlations reflect what a variable is not
  - Zeros provide definition by discriminant validity