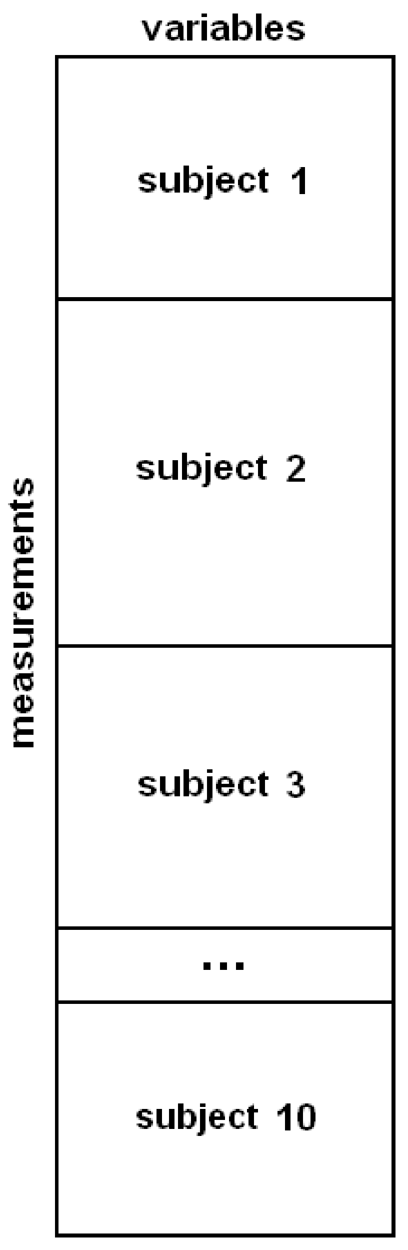
KATHOLIEKE UNIVERSITEIT
LEUVEN

# Clusterwise SCA-P

Kim De Roover, K.U.Leuven
Eva Ceulemans, K.U.Leuven
Marieke Timmerman, R.U.Groningen
Patrick Onghena, K.U.Leuven

COMPSTAT 2010

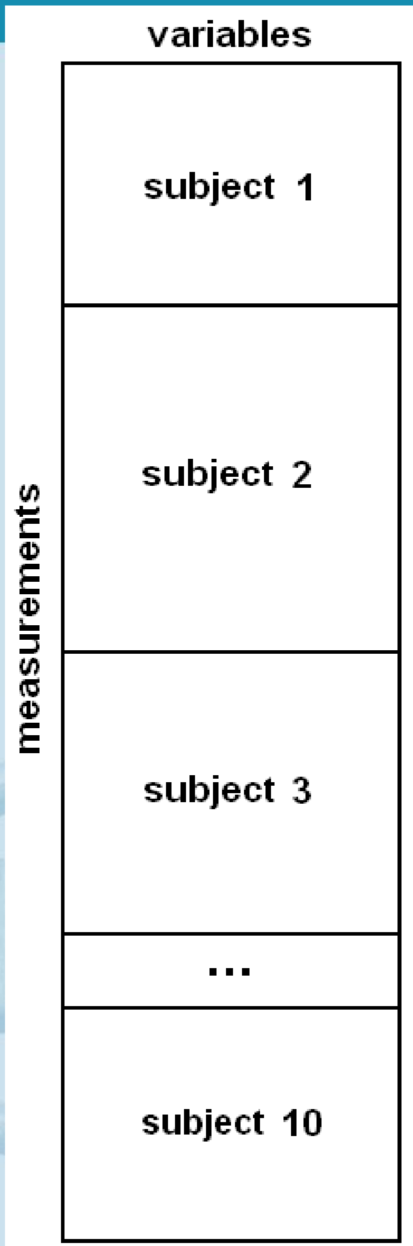
# Introduction




- data from different subjects with multiple measurements of a number of variables

➔ differences and similarities between subjects in underlying structure of the data?

# Illustrative application




Data (Vansteelandt et al., 2006):

- 10 subjects with eating disorder (anorexia and bulimia nervosa)
- 22 variables measuring: drive for thinness, positive and negative emotional states, urge to be physically active, physical activity
- 9 random measurement moments per day, during a week

Research questions:

(1) underlying structure of the variables?

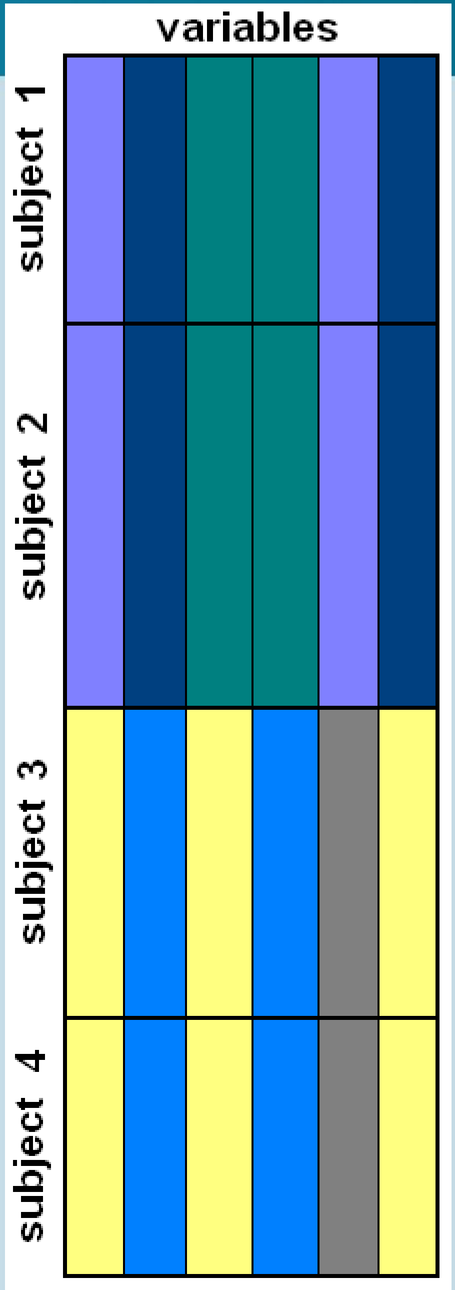
(2) interindividual differences in underlying structure?

# Clusterwise SCA: Idea

- general idea:

  - partition subjects into clusters

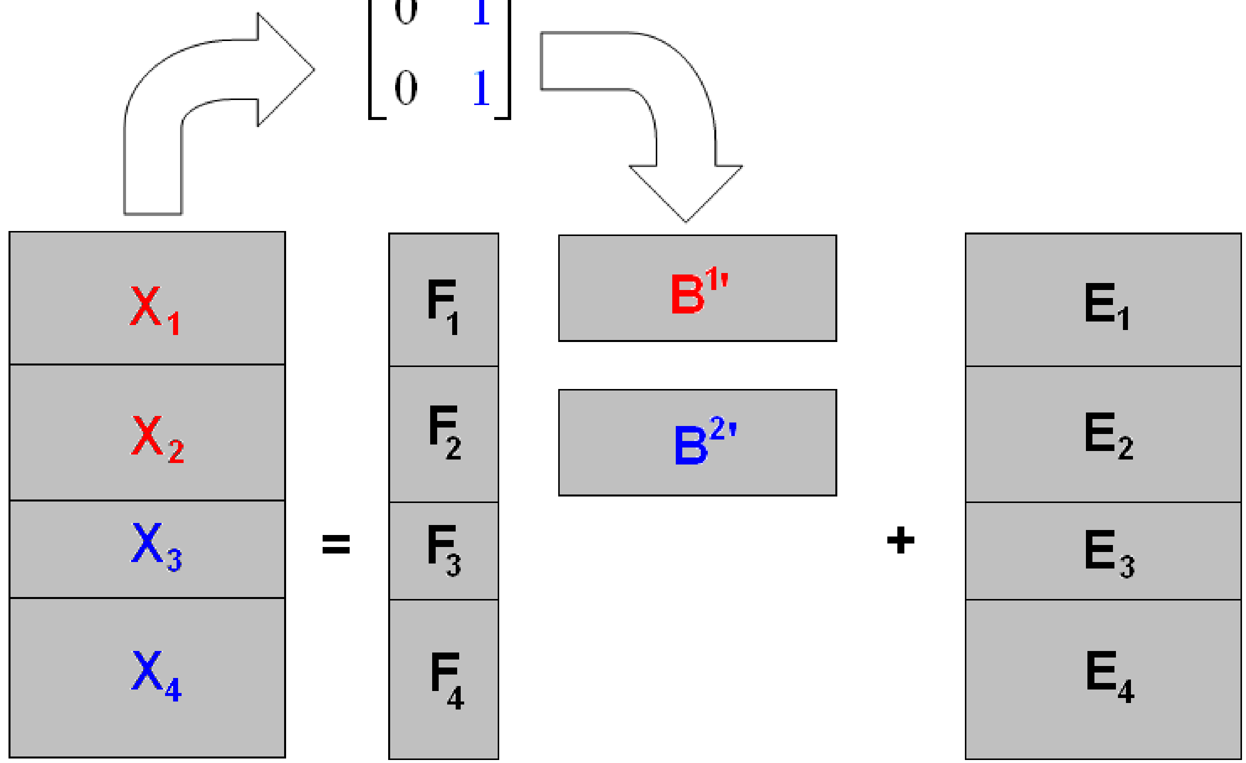
  - perform separate SCA per cluster



# Clusterwise SCA: Model

$$\mathbf{X}_i = \sum_{k=1}^{K} p_{ik}\mathbf{F}_i\mathbf{B}^{k}{}' + \mathbf{E}_i = \mathbf{F}_i\mathbf{B}^{k}_{i\in k}{}' + \mathbf{E}_i$$

$$\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$



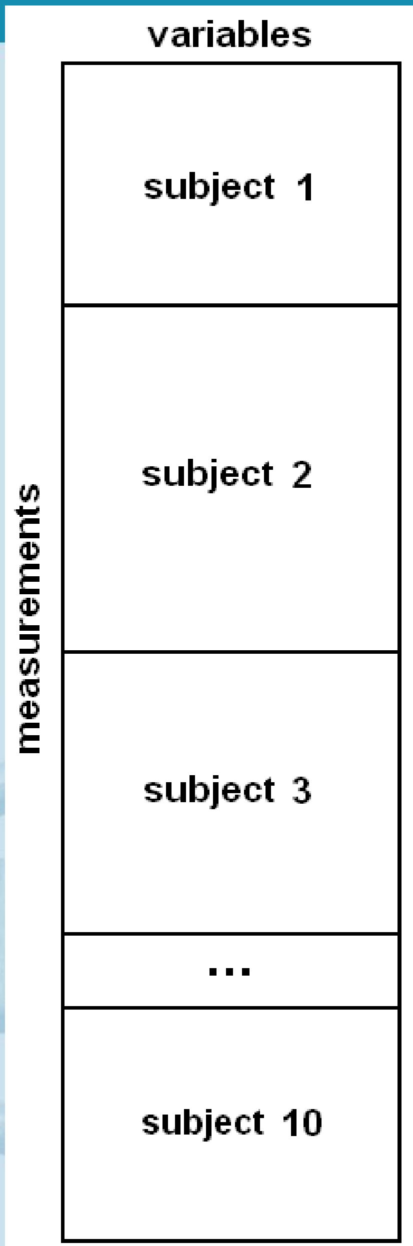
# Clusterwise SCA-P: Model

$$\mathbf{X}_i = \sum_{k=1}^{K} p_{ik} \mathbf{F}_i \mathbf{B}^{k}{}' + \mathbf{E}_i = \mathbf{F}_i \mathbf{B}^{k}_{i \in k}{}' + \mathbf{E}_i$$

- for now, number of components per cluster equal for all clusters

- more general than Clusterwise SCA-ECP (De Roover et al., 2010):
  - variances of components and correlations between components are allowed to vary between subjects within a cluster
    ➔ insight in differences between subjects in (co)variation of the components

# Illustrative application




Preprocessing of eating disorder data:

- centre per subject
  ➔ differences in mean scores between subjects removed

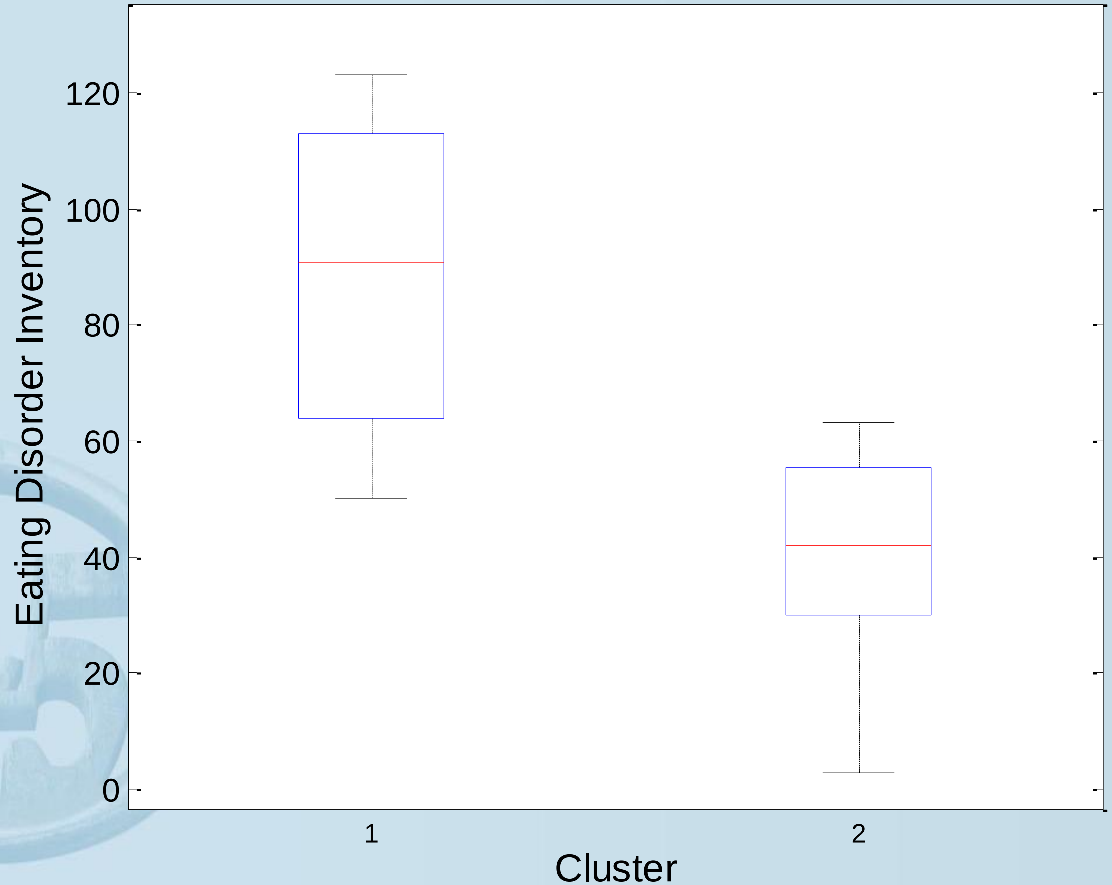
- standardize over the 10 subjects (instead of per subject)
  ➔differences in variability of the scores retained

# Illustrative application: Loadings

| *rotation criterion: HKIC (Harris & Kaiser, 1964)* | | **Cluster 1** (5 subjects) | | **Cluster 2** (5 subjects) | |
|---|---|---|---|---|---|
| | | *PA/NA* | *(urge) ph. act. & DFT* | *PA/NA* | *PA & ph.act.* |
| *positive and negative affect (PA/NA)* | pleased | **.80** | .19 | **.42** | **.47** |
| | happy | **.80** | .20 | **.32** | **.62** |
| | appreciated | **.61** | .14 | -.12 | **.76** |
| | love | **.56** | .23 | .01 | **.69** |
| | sad | **-.79** | -.17 | **-.77** | .08 |
| | angry | **-.79** | -.01 | **-.60** | .13 |
| | lonely | **-.71** | -.01 | **-.62** | -.02 |
| | ashamed | **-.68** | .05 | -.21 | .08 |
| | anxious | **-.75** | .03 | **-.41** | .03 |
| | tense | **-.68** | .10 | **-.49** | -.24 |
| | guilty | **-.81** | .05 | **-.44** | .00 |
| | irritated | **-.58** | .07 | **-.55** | -.07 |
| *urge to be physically active (urge ph. act.)* | want to move | -.08 | **.89** | .30 | -.23 |
| | want to sport | -.13 | **.94** | .11 | -.21 |
| | want to be active | -.07 | **.92** | .28 | -.27 |
| *physical activity (ph. act.)* | am active | .20 | **.87** | -.10 | **.55** |
| | am moving | .23 | **.87** | -.12 | **.59** |
| | am sporting | .29 | **.88** | .00 | .19 |
| *drive for thinness (DFT)* | want to burn calories | -.16 | **.89** | .10 | -.13 |
| | want to loose weight | **-.48** | **.32** | .02 | **-.43** |
| | feel fat | **-.66** | **.33** | -.02 | -.13 |
| | feel ugly | **-.60** | **.55** | -.02 | -.20 |

# Illustrative application: Validation of clustering

significant difference between clusters in mean EDI (p = .04)



# Illustrative application: Variances/correlations per cluster

| | subject | variances PA/NA | variances (urge) ph. act. & DFT | correlations |
|---|---|---|---|---|
| **cluster 1** | 1 | .91 | .82 | .17 |
| | 2 | 1.01 | .71 | -.28 |
| | 3 | .89 | 1.31 | -.18 |
| | 4 | 1.20 | 1.11 | -.05 |
| | 10 | 1.03 | .81 | .05 |
| | | | ***overall correlation cluster 1*** | ***-.05*** |
| | *subject* | *variances PA/NA* | *variances PA & physical activity* | *correlations* |
| **cluster 2** | 5 | .51 | 1.45 | -.52 |
| | 6 | 1.56 | 1.31 | .45 |
| | 7 | .46 | .36 | .49 |
| | 8 | 1.08 | 1.34 | .45 |
| | 9 | 1.59 | .66 | .12 |
| | | | ***overall correlation cluster 2*** | ***.21*** |

# Illustrative application: Variances/correlations per cluster

.39 correlation with EDI

-.11 correlation with EDI

.99 correlation with EDI

| | *subject* | *variances PA/NA* | *variances (urge) ph. act. & DFT* | *correlations* |
|---|---|---|---|---|
| **cluster 1** | 1 | .91 | .82 | .17 |
| | 2 | 1.01 | .71 | -.28 |
| | 3 | .89 | 1.31 | -.18 |
| | 4 | 1.20 | 1.11 | -.05 |
| | 10 | 1.03 | .81 | .05 |
| | | | ***overall correlation cluster 1*** | ***-.05*** |

| | *subject* | *variances PA/NA* | *variances PA & physical activity* | *correlations* |
|---|---|---|---|---|
| **cluster 2** | 5 | .51 | 1.45 | -.52 |
| | 6 | 1.56 | 1.31 | .45 |
| | 7 | .46 | .36 | .49 |
| | 8 | 1.08 | 1.34 | .45 |
| | 9 | 1.59 | .66 | .12 |
| | | | ***overall correlation cluster 2*** | ***.21*** |

.21 correlation with EDI

.89 correlation with EDI

-.60 correlation with EDI

# Discussion

Clusterwise SCA-P:

- captures structural differences and similarities in a parsimonious manner
- makes it possible to examine differences in component variances and correlations between the subjects within a cluster
- is applicable to all kinds of multivariate nested data, e.g., subjects nested within groups
- number of components will be allowed to vary over clusters in the future