UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

Jumpstart Mplus

1. Introduction to SEM

Dr Gabriela Roman

Arielle Bonneville-Roussy

UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

What is SEM?

- Structural = characterised by structure
- Equation = relation between variables
- Model = pattern, representation

- SEM = a statistical technique for testing and estimating causal relationships using statistical data and causal assumption *(Wikipedia)*

SEM

- SEM process:
  - →Formulate theoretical model
  - →Apply model to the data
  - →Check whether the model fits the data
  - →Examine the relationships
  - →Refine the model (repeat until satisfied)

- Model components:
  - Model fit
  - Estimates

SEM – Model fit indices

Indicator of fit	Fit is good if…	Notes:
Chi-square	Should be non-sig.	It almost always is sig. when you have a large sample!
CFI (Comparative fit index)	> .95 = good > .90 = 'adequate'	
TLI (Tucker-Lewis index)	> .95 = good > .90 = 'adequate'	Penalises overly complex models.
RMSEA (Root mean square error of approximation)	< .06 = good	Penalises overly complex models. Comes with a 90% confidence interval.

measures simply whether the model specified a priori successfully reproduces data patterns

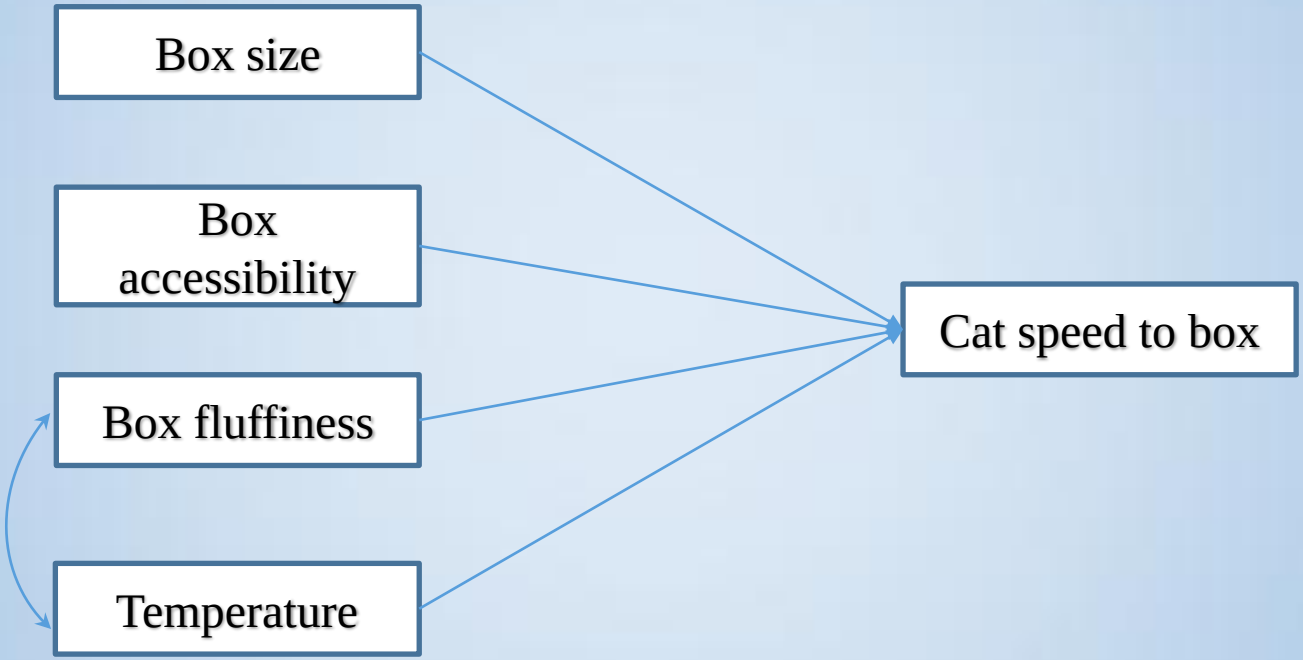
measure whether the hypothesised model fits the data better than a more restricted baseline model

Study

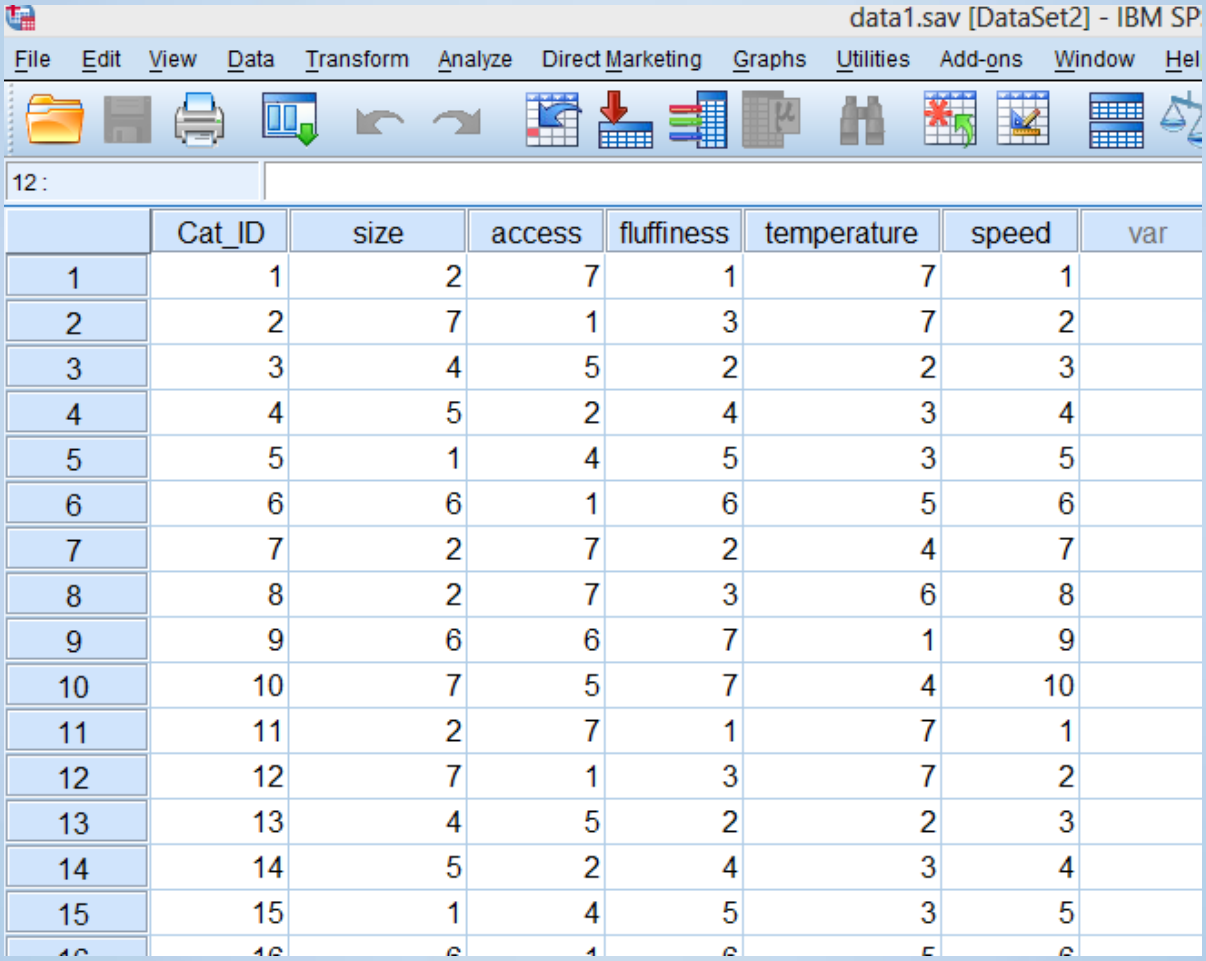
- What predicts a cat's speed toward the box:
  - Box size (way too small – huge)
  - Box accessibility (access is impossible – easy access)
  - Degree of box "fluffiness" (not fluffy – very fluffy)
  - Weather conditions (temperature: freezing – hot)

- Box fluffiness and weather conditions may be slightly correlated, assuming the cat owns a caring human to add a blanket during the cold season.

Study

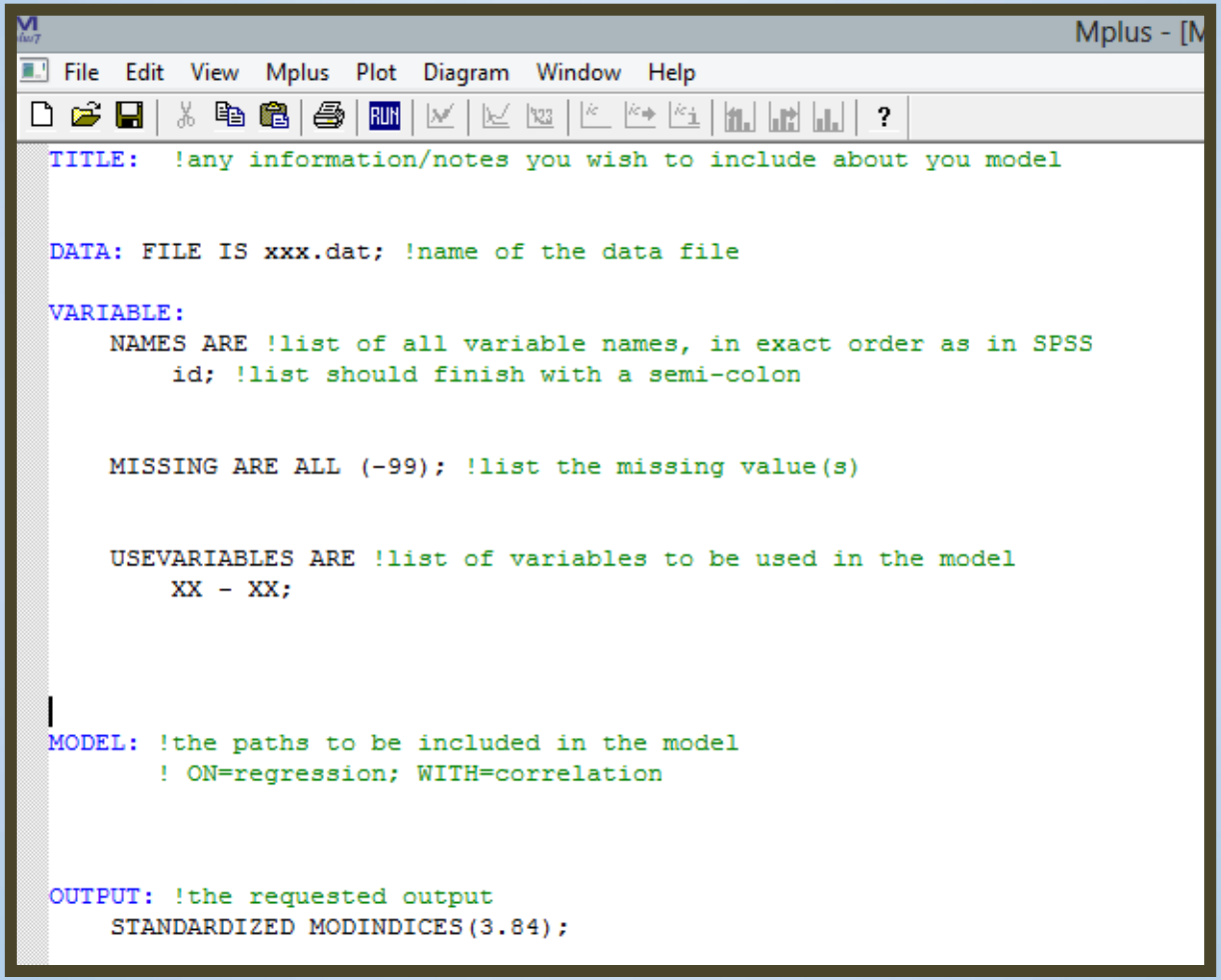

Box size
Box accessibility
Box fluffiness
Temperature
Cat speed to box

Preparing data for Mplus

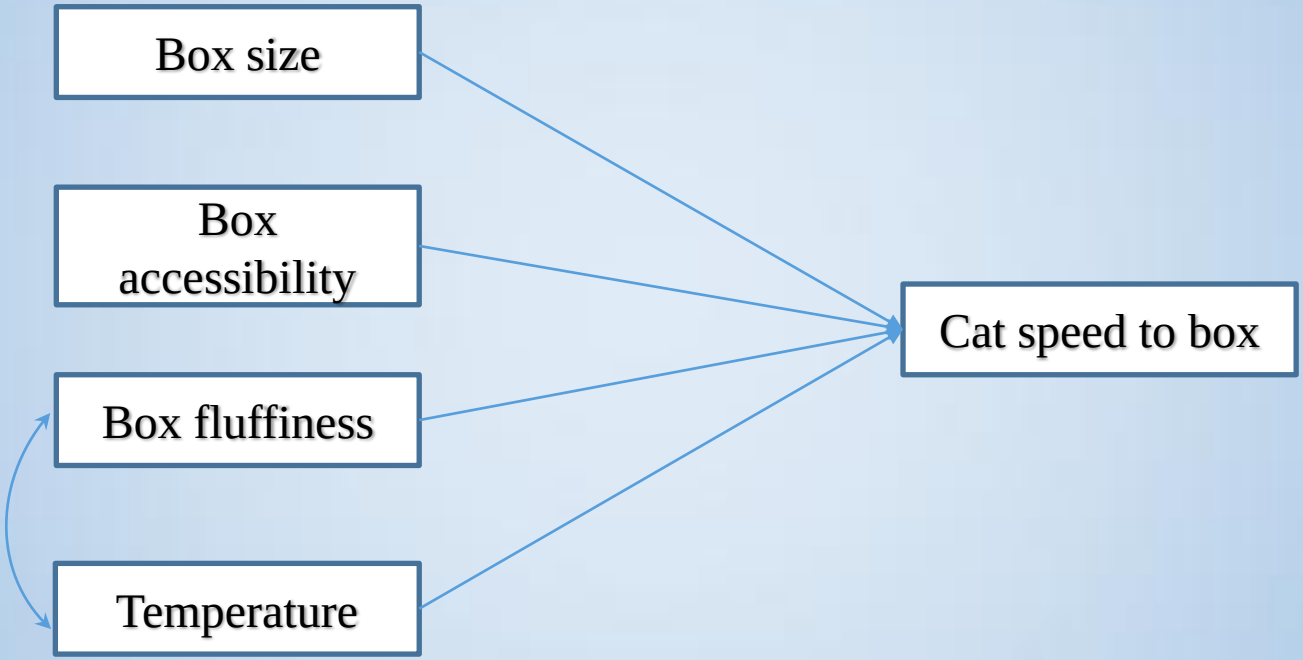


	Cat_ID	size	access	fluffiness	temperature	speed	var
1	1	2	7	1	7	1	
2	2	7	1	3	7	2	
3	3	4	5	2	2	3	
4	4	5	2	4	3	4	
5	5	1	4	5	3	5	
6	6	6	1	6	5	6	
7	7	2	7	2	4	7	
8	8	2	7	3	6	8	
9	9	6	6	7	1	9	
10	10	7	5	7	4	10	
11	11	2	7	1	7	1	
12	12	7	1	3	7	2	
13	13	4	5	2	2	3	
14	14	5	2	4	3	4	
15	15	1	4	5	3	5	

Mplus input file



Study


Box size
Box accessibility
Box fluffiness
Temperature
Cat speed to box

Mplus - [M

File Edit View Mplus Plot Diagram Window Help

```
TITLE:   !any information/notes you wish to include about you model

DATA: FILE IS xxx.dat; !name of the data file

VARIABLE:
    NAMES ARE !list of all variable names, in exact order as in
        id; !list should finish with a semi-colon

    MISSING ARE ALL (-99); !list the missing value(s)

    USEVARIABLES ARE !list of variables to be used in the model
        XX - XX;

MODEL: !the paths to be included in the model
       ! ON=regression; WITH=correlation

OUTPUT: !the requested output
    STANDARDIZED MODINDICES(3.84);
```

File Edit View Mplus Plot Diagram Window Help

```
TITLE:

DATA: FILE IS cat-in-box.dat;

VARIABLE:
    NAMES ARE
        cat_id size access fluffy temp speed;

    MISSING ARE ALL (-99);

    USEVARIABLES ARE
       size access fluffy temp
       speed;

MODEL:

    speed ON size;
    speed ON access;
    speed ON fluffy;
    speed ON temp;

    fluffy WITH temp;

OUTPUT:
    STANDARDIZED MODINDICES(3.84);
```

Results?

UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

Jumpstart Mplus

2. Introduction to factor analysis

Arielle Bonneville-Roussy

Dr Gabriela Roman

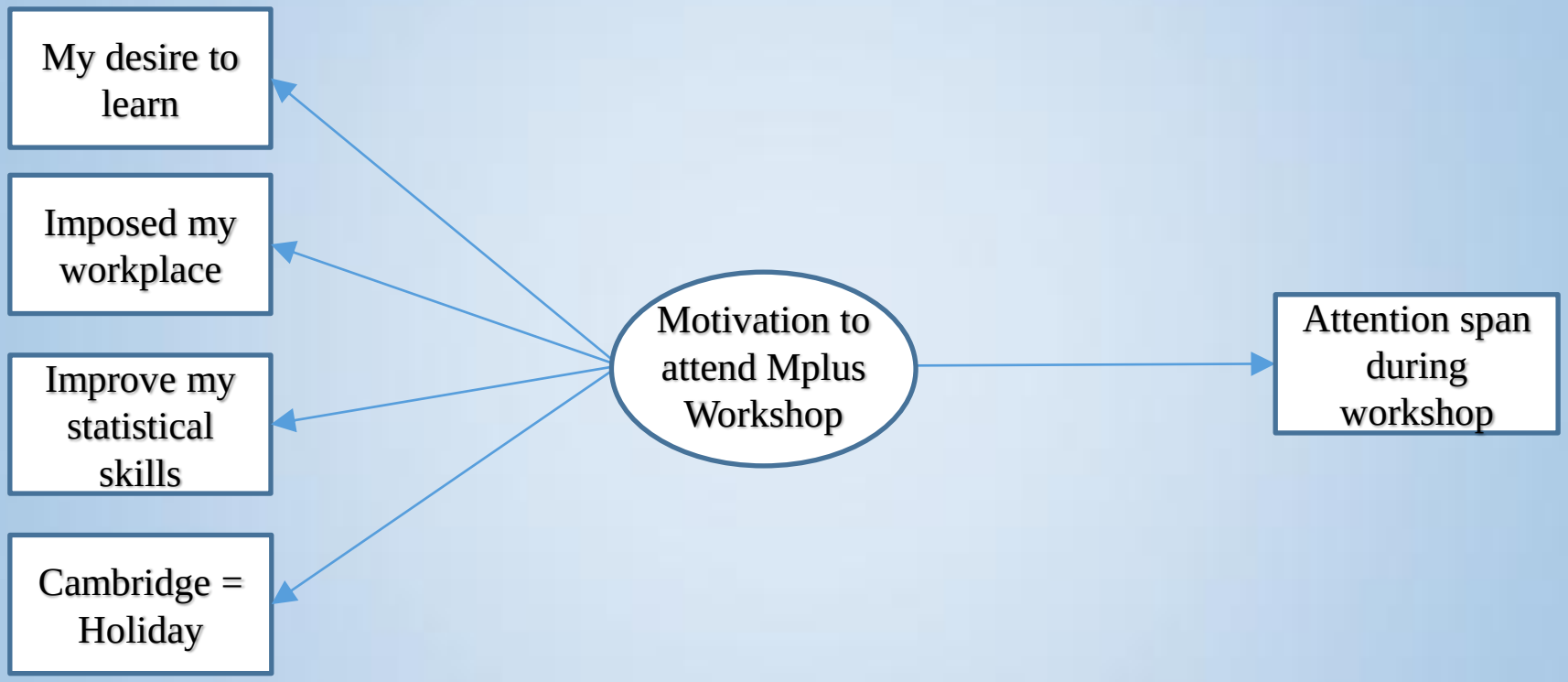
UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

Questions

- What are measurement models?
- Why do we use measurement models?
  - The Common Factor Model
- How can we apply measurement models?
  - The Principles of Factor Analysis
  - The Differences between EFA and CFA
  - The Assumptions of EFA and CFA
  - The Model Specifications
  - Mplus Outputs Related to Measurement Models

Regression vs Measurement models



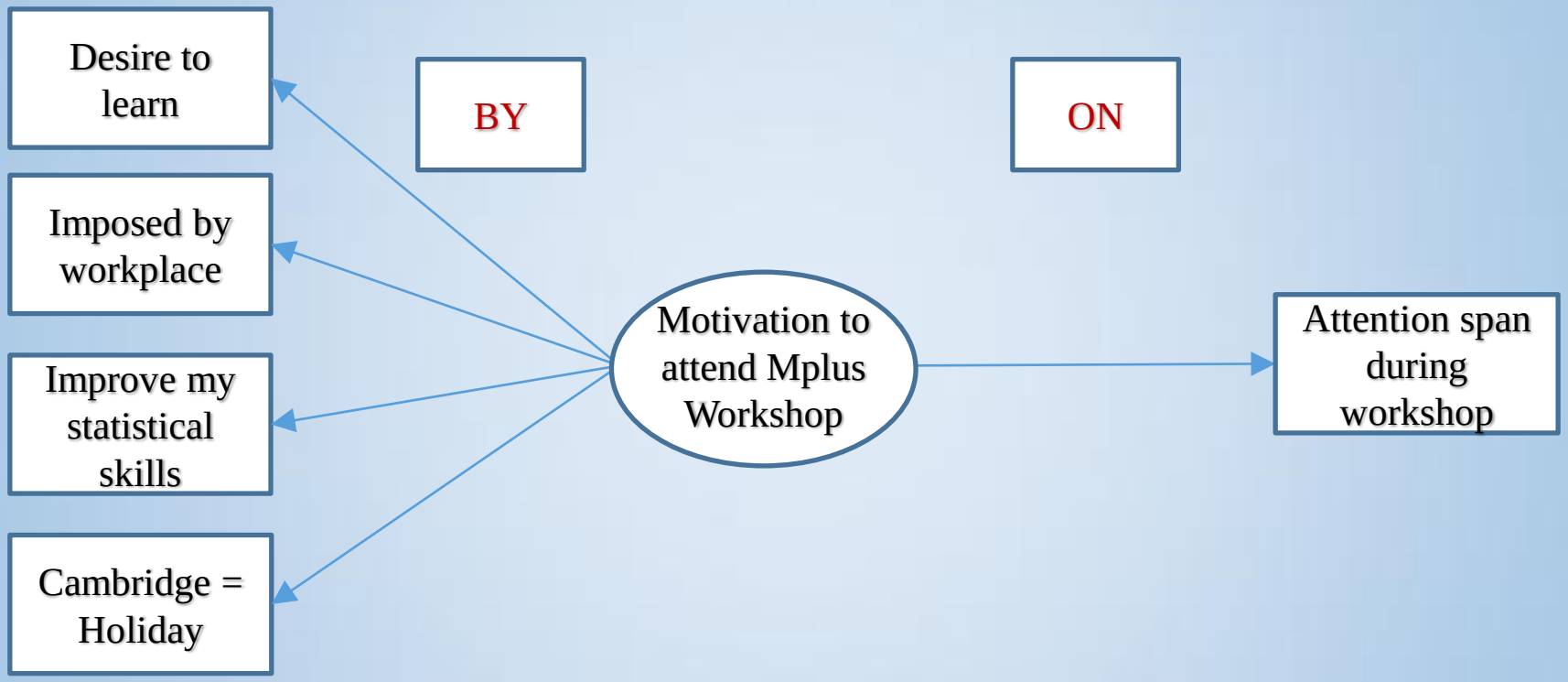
Regression vs Measurement models

- In SEM, measurement models (as opposed to regression models) are the links between observed and latent variables.
- Regression Models:
  - Cats in the box
  - Regression paths
  - Path analysis
  - Mplus ON statement

Regression vs Measurement models

- Measurement models:
  - Exploratory Factor Analysis (EFA)
  - Confirmatory Factor Analysis (CFA)
  - Latent class
  - Mplus BY statement
- Regression and measurement models combined:
  - Structural Equation Modelling (SEM)

Regression vs Measurement models



What are measurement models?

- The most common measurement model:

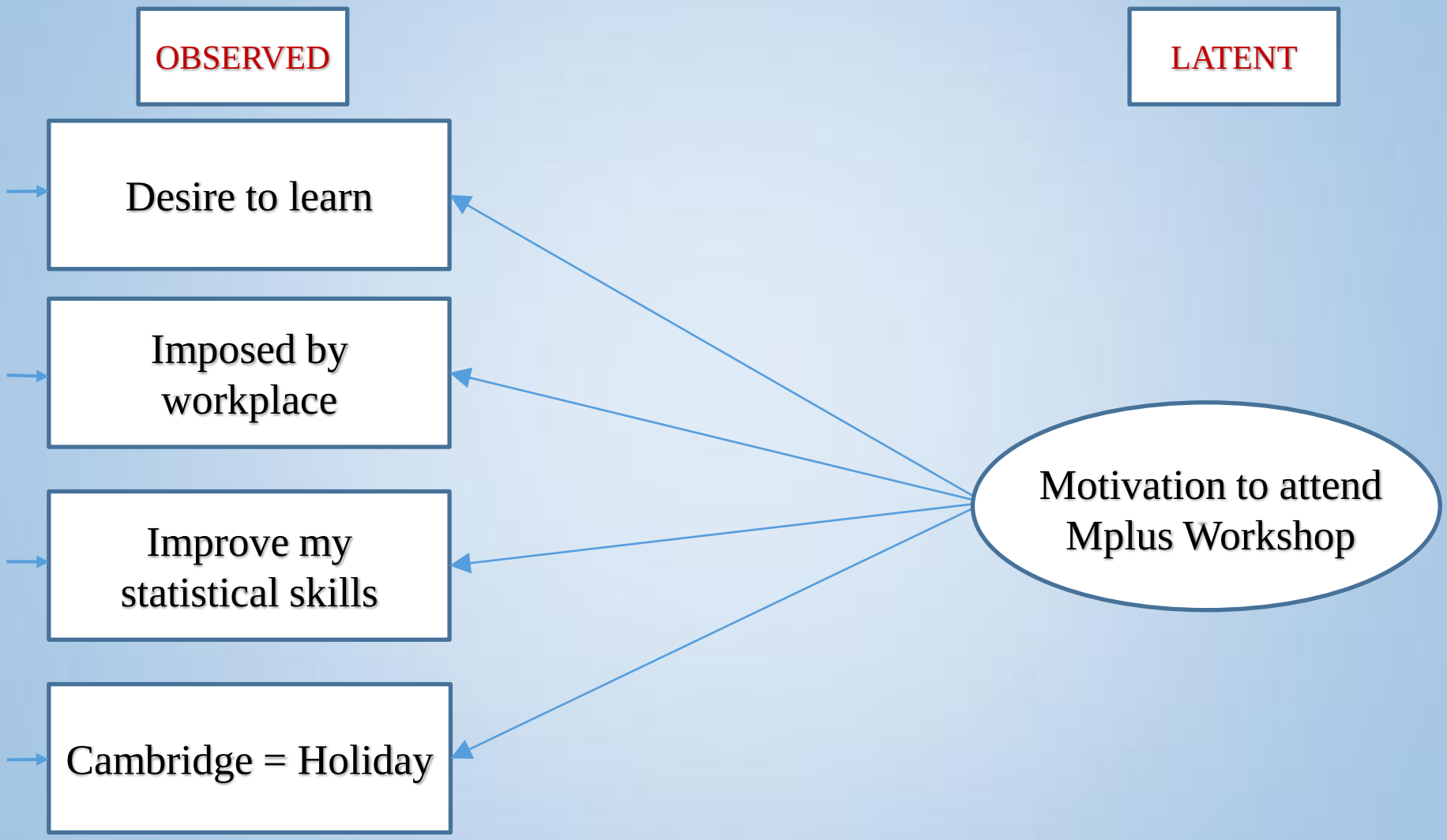
  Common Factor Model

- Some observed measures share a common cause/construct: they are intercorrelated
- E.g.: Sleeping problems and lack of appetite in depression

What are measurement models?

- Overarching goal :
  - To measure latent variables by relating them with things that can be observed
- Observed variables (indicators):
  - Variables that are either observed or assessed directly
  - Ex: age; heart rate, response to a single item on a Likert scale
- Latent variables (factors):
  - Variables (phenomena) that are not measured or observed directly.
  - Ex: personality, depression, intelligence, physical activity by self-report

Measurement models



What are measurement models?

- Factor analysis (FA)
  - Assesses the dimensionality of a set of observed variables
- Dimensions = Factors
  - Dimensions/factors/latent variables are underlying hypothetical constructs
- Factor =
  - Common variance between the observed variables
  - Unique variance of each of the observed variable
  - Measurement error

Why do we use measurement models?

- To reduce or to find dimensionality
- For testing and scale construction
- Validity of a measure
- To assess a theoretical construct
- To find patterns of correlations between a set latent variables
- To score participants on latent variables

Why do we use measurement models?

With a fully validated questionnaire

- To account for weight (factor loadings) when assessing a phenomenon
  - For instance:
  - In depression, which one should have more weight:
    - Suicidal thoughts?
    - Irritability?

Factor Analysis

- Exploratory and confirmatory factor analysis
- Both are methods of data reduction
- Aim to replicate the relationships (correlations) between a set of observed variables

Factor Analysis Equation

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Factor Analysis

General Equation

$$y = \Lambda_y \eta + \varepsilon$$

Observed Variables = Factors + Error

- Where:
  - y = a set of observed variables
  - n = factors
  - ε = and residuals (error)
  - Λ = matrix

EFA VS CFA

	EFA	CFA
Aim	To explore dimensionality by finding the smallest number of factors that explains the observed variables	To confirm the validity of a pre-determined factor structure
Foundation	Data driven	Theory driven
Number of factors	No initial specification	Specified
Type of matrix analysed	Correlation	Variance/covariance
Type of solution	Standardised	Standardised/Unstandardised
Structure	Obtained with factor rotation	Obtained by fixing cross-loadings to zero

Assumptions of EFA and CFA

- For a straightforward analysis, data are:
  - Continuous (pseudo-continuous)
  - Normally distributed
  - No missing data or missing at random (by design)
  - Large enough sample size
  - Absence of multicollinearity

- Otherwise: Wait for tomorrow morning....

Continuous data

- Pure continuous
  - Age, height
- Pseudo continuous
  - Likert scales > 4 are sometimes considered continuous AND approaching normal distribution
  - > 9 points are considered by most as continuous

Normally distributed

- Absence of substantial skewness or kurtosis in one or more variable(s)
- Skewness and Kurosis tests are available in SPSS, R etc.
- Otherwise consider transformation of variables

No missing data or missing at random (by design)

- Absence of missing data
- Missing data are at random:
  - E.g. To minimise fatigue, participants answer to a randomly chosen subset of 50% of the items
- MCAR and MAR tests available in SPSS, R, Mplus etc.
- MCAR missing data are dealt with in Mplus

Large enough sample size

- Brown (2006), Chapter 10
  - P. 419 Satorra-Saris method to calculate sample size
- Rule of thumb?
  - N:*q* rule (Jackson, 2003; Kline, 2011)
  - N= sample size, *q* = number of parameters
    - 20:1
  - > 200 or 10 cases per observed variable?
- Consequences of a small sample size:
  - Sometimes: none
  - Large standard errors (confidence interval)
  - Solution does not converge
  - Low statistical power

Absence of extreme multicollinearity (for EFA)

- Two or more observed variables are highly correlated. e.g. r > .90
- Screening of the correlation matrix
- Tolerance and VIF tests in SPSS, R, etc.
- Solution: remove one of the variables, obtain more data...

Model Specifications

- Estimators/ Fitting procedure/Factor extraction technique (EFA and CFA)
  - They calculate a set of orthogonal components that are sought to best reproduce the relationships between the variables
    - ML
    - ULS
    - WLS
    - MUML

Model Specifications

- Rotation (EFA)
- Improves the interpretability of the solution
  - Orthogonal: Estimated factors are uncorrelated
    - Useful when the factors are sought to be completely independent
    - Rare in biological/social sciences
  - Oblique: Estimated factors are correlated
    - Often the case in psychology

Factor loadings

- Weight (importance) of the variable to the factor
- Unstandardised (CFA)
  - Loadings can be greater than 1
  - Based or the original metric
  - For instance heart rate (OV) and passion (LV)
    - One unit increase in passion corresponds to an increase in heart rate of 15 beats per minute
- Standardised (EFA and CFA)
  - Between 0 and 1
  - To be interpreted as a regression coefficient
    - < .30-40 = low explanatory power
    - .40-.70 = medium explanatory power
    - > . 70 = high explanatory power

Summary

- Measurement models in SEM are the components that assess the dimensionality of a set of observed variables.
- They are mostly types of factor analyses.
- Those covered in this course: EFA and CFA.
- They both have assumptions that if met, make your life easier.
- Both CFA and EFA need to be specified in several ways for convergence.

Suggested steps for the validation of a questionnaire

1. Item selection or generation

1. Pre-test with a sample of at least n +1 , EFA
   - Purification of the questionnaire
   - Remove items with poor loadings, with multicollinearity, etc...
   - CFA, measurement invariance, parallel tests

Suggested steps for the validation of a questionnaire

3. Preliminary study with a larger sample (see guidelines in Brown, 2006; Kline, 2010), EFA
   - Determine the number of factors

4. Main study (studies), sample size representative of the population, CFA
   - CFA, measurement invariance, parallel tests

UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

Jumpstart Mplus

3. Exploratory Factor Analysis

Arielle Bonneville-Roussy

Dr Gabriela Roman

UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

Questions

- What is the purpose of EFA?
- What are the characteristics of EFA?
- How to perform EFA in Mplus?
- How to interpret EFA?

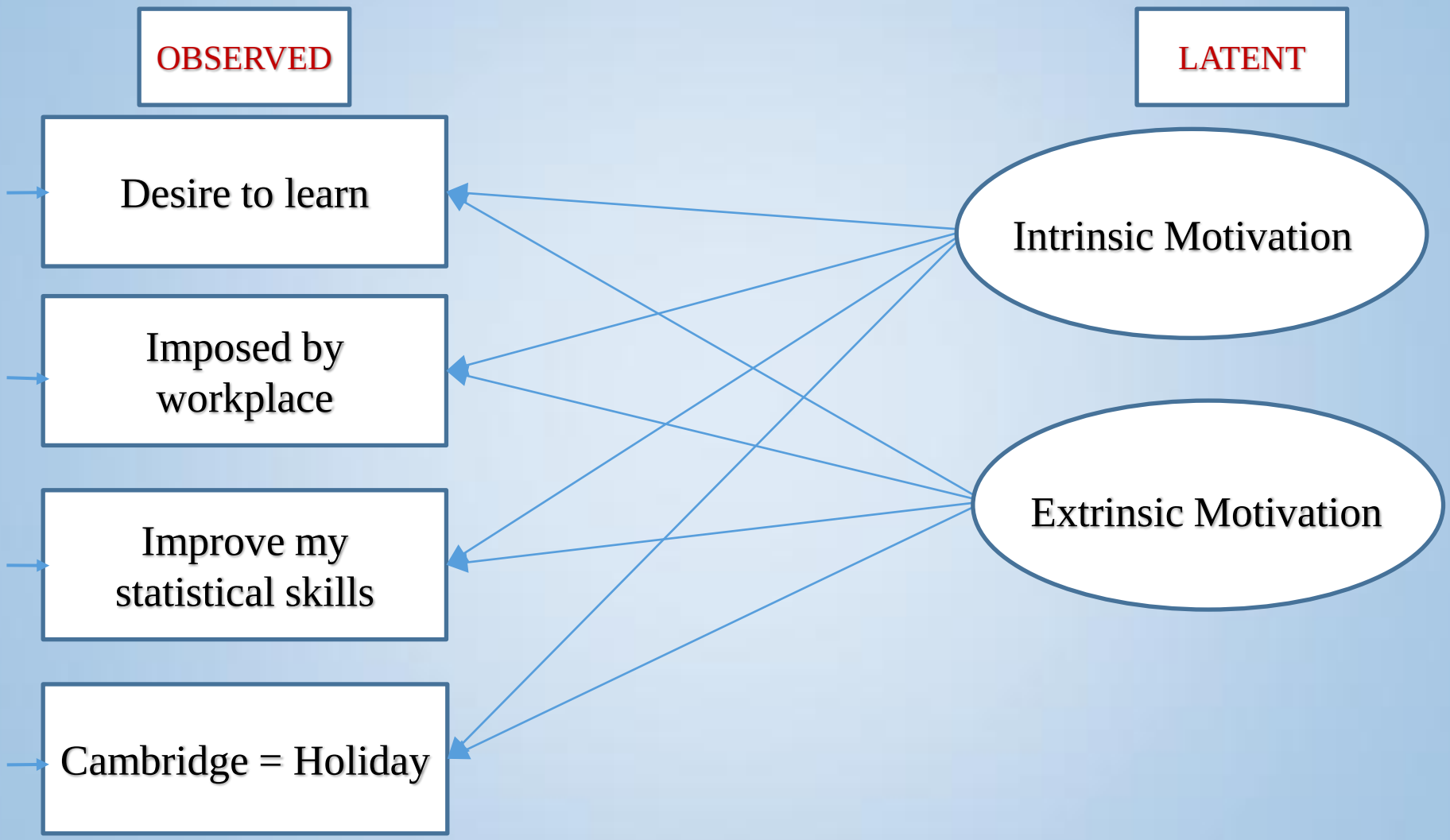
What is the purpose of EFA?

- To explore preliminary data
- To find the smallest number of factors
- To screen for observed variables that poorly explain a theoretical construct
- To screen for two or more observed variables that can equally explain a theoretical construct (principle of parsimony)

What are the characteristics of EFA?

- All of the observed variables are related to every factor
- Uses a correlation matrix
  - Solutions are always standardised
- Data driven
- EFA solutions are usually rotated
  - Factors can be correlated or not

EFA

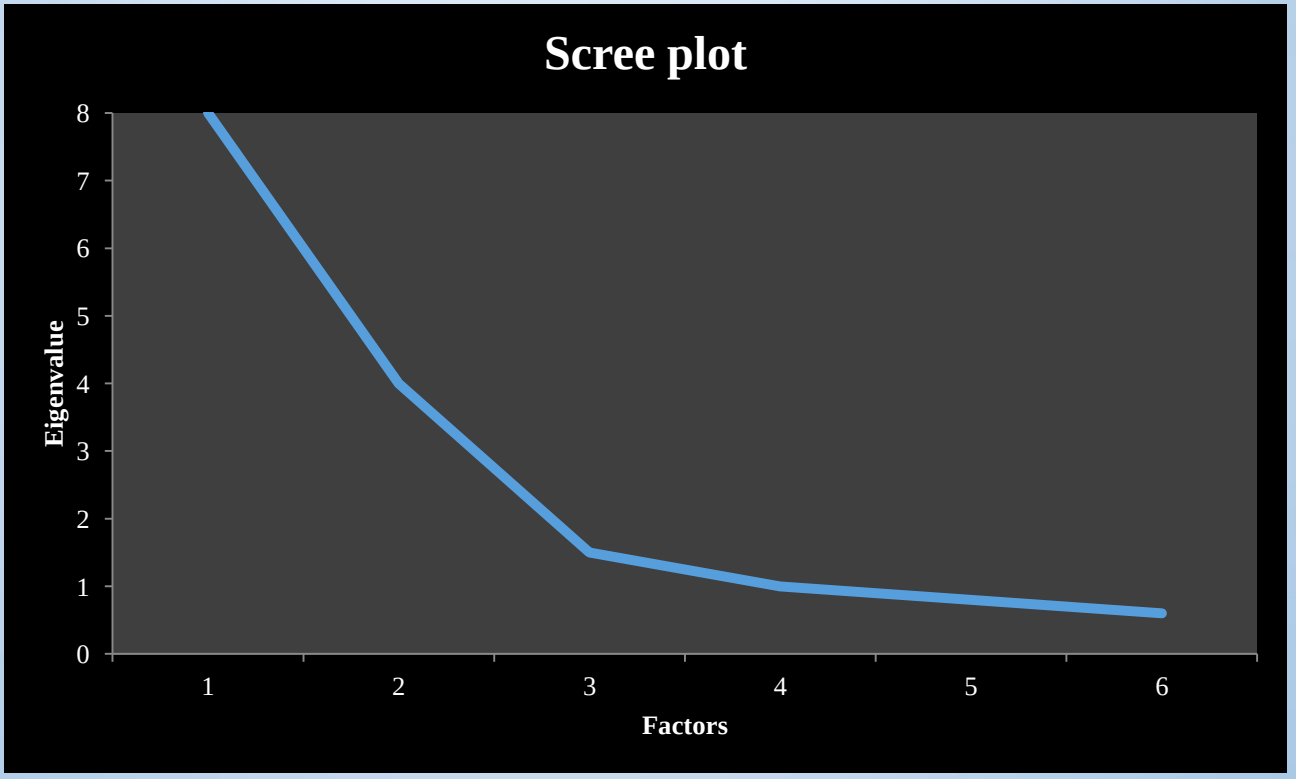

OBSERVED
LATENT
Desire to learn
Imposed by workplace
Improve my statistical skills
Cambridge = Holiday
Intrinsic Motivation
Extrinsic Motivation

Determining the numbers of factor

1. Eigenvalue
   - Sum of the squared factor loadings for each factor
   - Gives us how much variance is explained by the observed variables for each factor
   - Criterion > 1

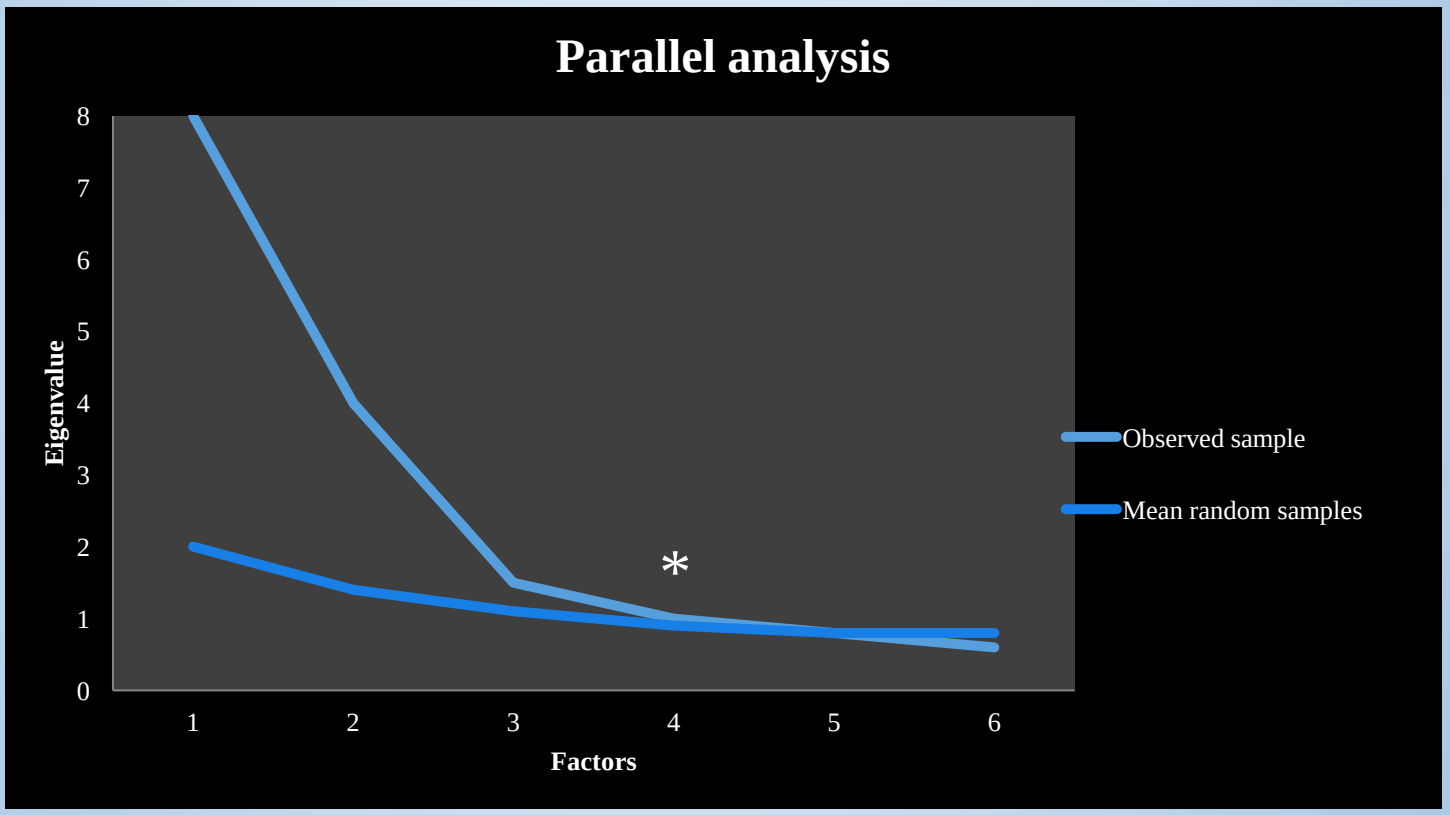
Determining the numbers of factor

2. Scree plot
   - We look for the elbow



Determining the numbers of factor

3. Parallel analysis



Determining the numbers of factor

4. Fit indices

Indicator of fit	Fit is good if…	Notes:
Chi-square	Should be non-sig with a small sample size	It almost always is significant when you have a large sample!
CFI (Comparative fit index)	> .95 = good > .90 = 'adequate'	
TLI (Tucker-Lewis index)	> .95 = good > .90 = 'adequate'	Penalises overly complex models.
RMSEA (Root mean square error of approximation)	< .06 = good	Penalises overly complex models. Comes with a 90% confidence interval.

Rotation

OBLIQUE vs orthogonal

Oblique = correlated

Mplus:

- GEOMIN = Default
- QUARTIMIN
- PROMAX
- Other methods

Rotation

OBLIQUE vs orthogonal

Orthogonal = uncorrelated

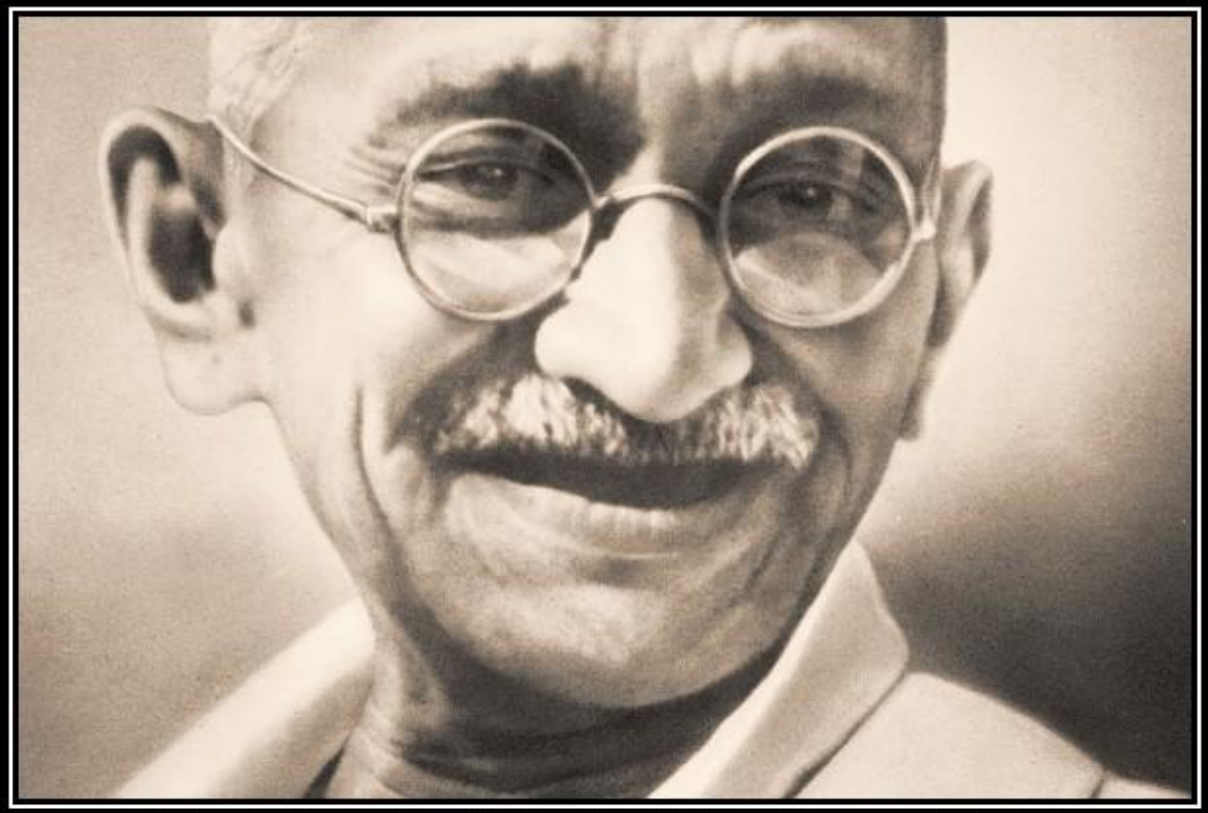
Mplus:

- GEOMIN (orthogonal)
- QUARTIMIN (orthogonal)
- VARIMAX

Mplus

Example

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre



MOTIVATION

"Whatever you do will be insignificant, but it is very important that you do it."

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Mplus Example

The study

Motivation to attend an Mplus Course

Derived from Global Motivation Scale (Guay et al., 2003), 7-point likert scale.

I attend an Mplus course…

- EXT …in order to show others what I am capable of. SHOW
- INT ... because I like making interesting discoveries. INTER
- INT ... because of the pleasure I feel as I become more and more skilled. SKILL
- EXT ... because I do not want to disappoint certain people. PLEASE
- EXT ... because I want to be viewed more positively by certain people. POSIVIEW
- INT ... because of the sense of well-being I feel while I am doing them. WELL

Mplus Example

```
title:   Exploratory factor analysis
data:   file is "jumpstartmplus.dat";


VARIABLE:

NAMES are
gender age show inter skill please posiview well;

USEVARIABLES  are
show inter skill please posiview well;

ANALYSIS:

!TYPE = BASIC ; ! To examine means and correlations

TYPE = EFA 1 3; ! Calls exploratory factor analysis with a minimum of 1 factor and a maximum of 3 factors

ESTIMATOR= ML; ! Default
ROTATION= GEOMIN;! Default; Other rotation procedures include Oblimin, Varimax, Promax...

parallel (50); ! Draws 50 random samples to be compared to the observed data

PLOT:

TYPE = PLOT2; !Eigenvalues (scree) plot
```

UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

Jumpstart Mplus

4. Confirmatory Factor Analysis

Dr Gabriela Roman

Arielle Bonneville-Roussy

UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

What is CFA?

- Confirmatory = structure is known (imposed by researcher)
- Factor = latent, unobserved trait
- Analysis = in this case: method of data reduction

- CFA = a method of data reduction used to examine the presence of an unobserved trait that explains the common variance of observed indicators

CFA

	EFA	CFA
Aim	To explore dimensionality by finding the smallest number of factors that explains the observed variables	To confirm the validity of a pre-determined factor structure
Foundation	Data driven	Theory driven
Number of factors	No initial specification	Specified
Type of matrix analysed	Correlation	Variance/covariance
Type of solution	Standardised	Standardised/Unstandardised
Structure	Obtained with factor rotation	Obtained by fixing cross-loadings to zero

Study

- Personal growth initiative:

  = a person's active an intentional involvement in changing and developing as a person.

1	2	3	4	5	6
Definitely disagree	Mostly disagree	Somewhat disagree	Somewhat agree	Mostly agree	Definitely agree

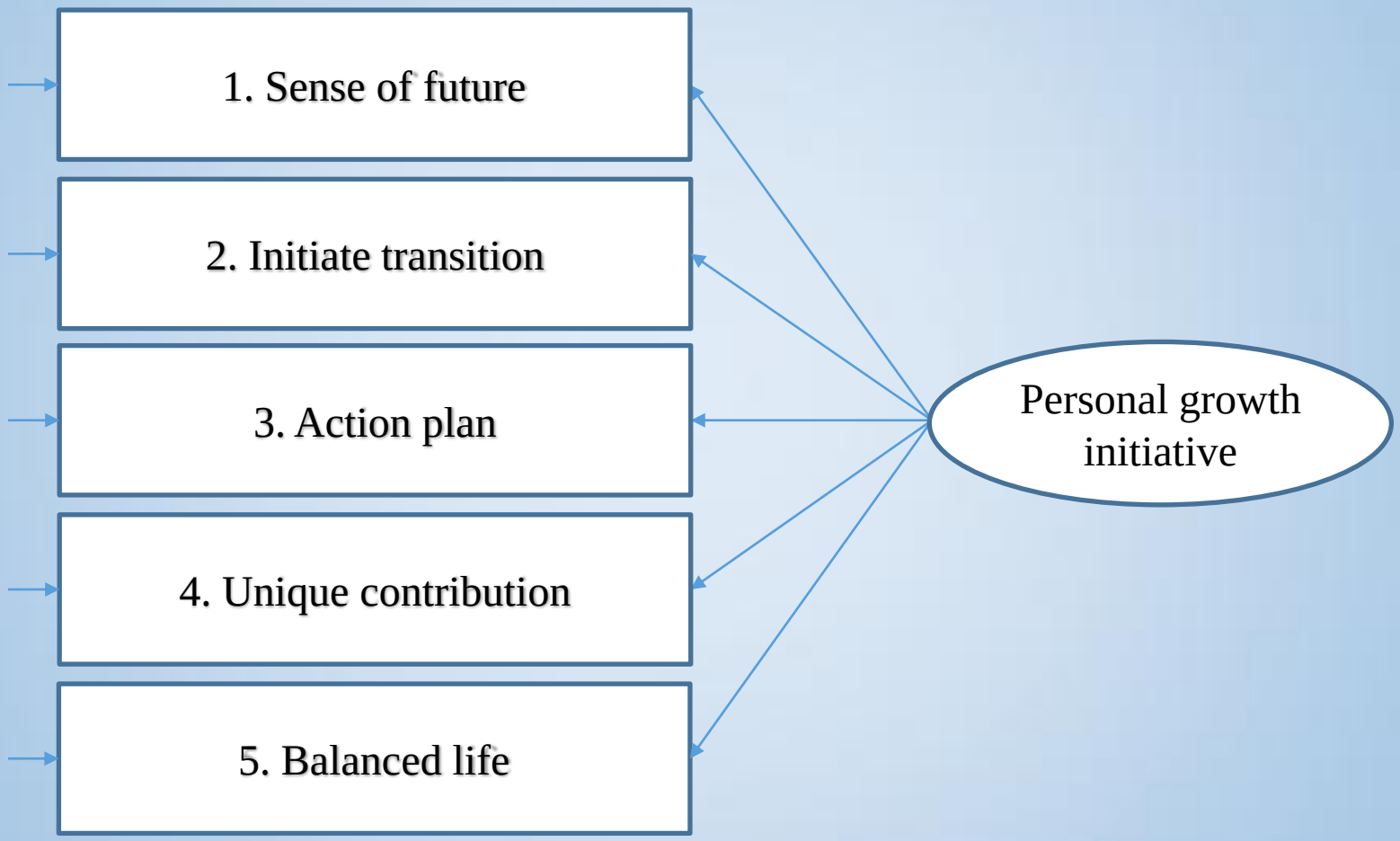
Personal growth initiative scale

- Questions: *(Robitschek, 1998)*

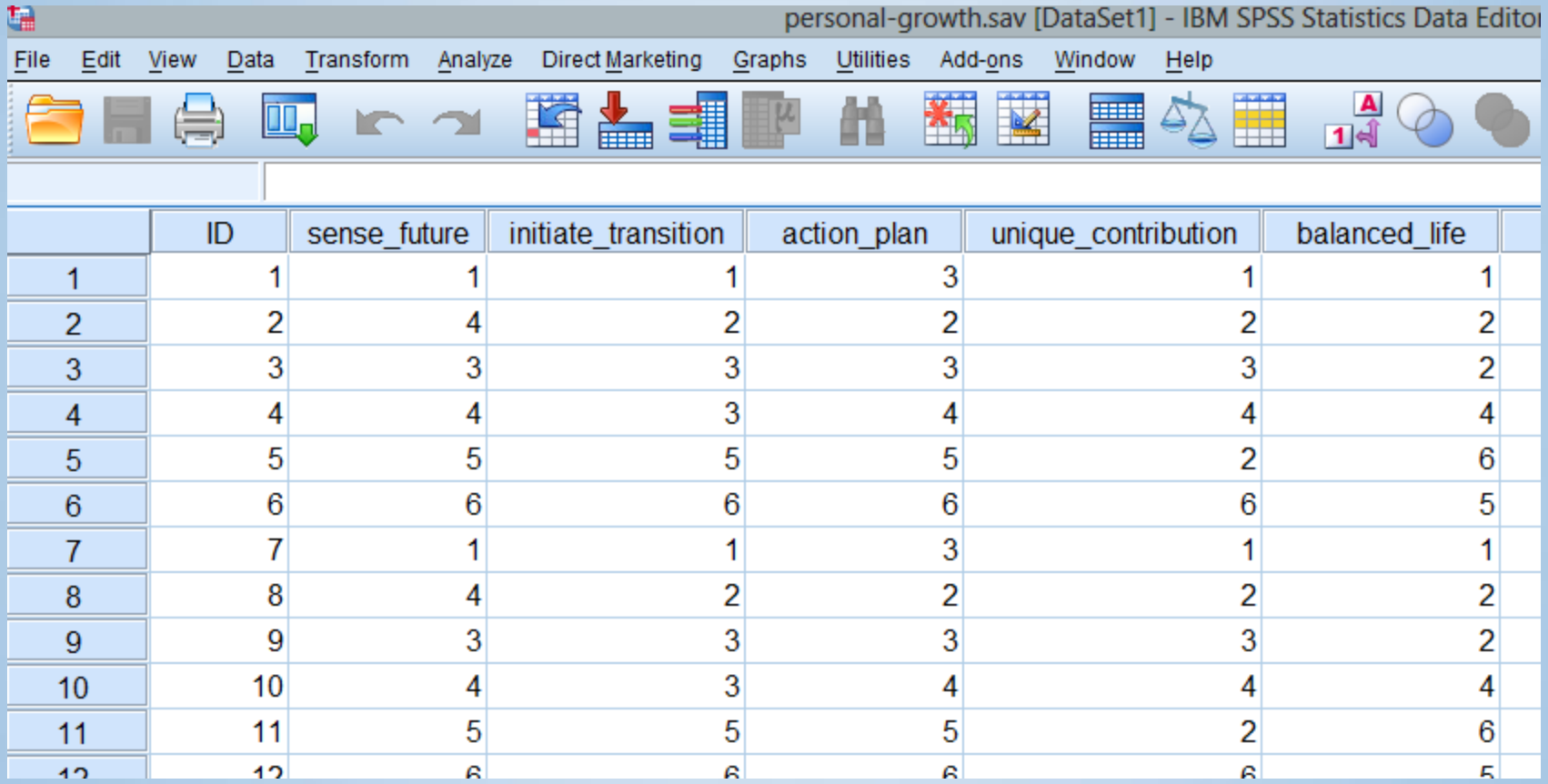
1. I have a good sense of where I'm headed in my life.
2. If I want to change something in my life, I initiate the transition process.
3. I have a specific action plan to help me reach my goals.
4. I know what my unique contribution to the world might be.
5. I have a plan for making my life more balanced.

**Note: the full questionnaire contains 9 items.*

Study


1. Sense of future
2. Initiate transition
3. Action plan
4. Unique contribution
5. Balanced life
Personal growth initiative

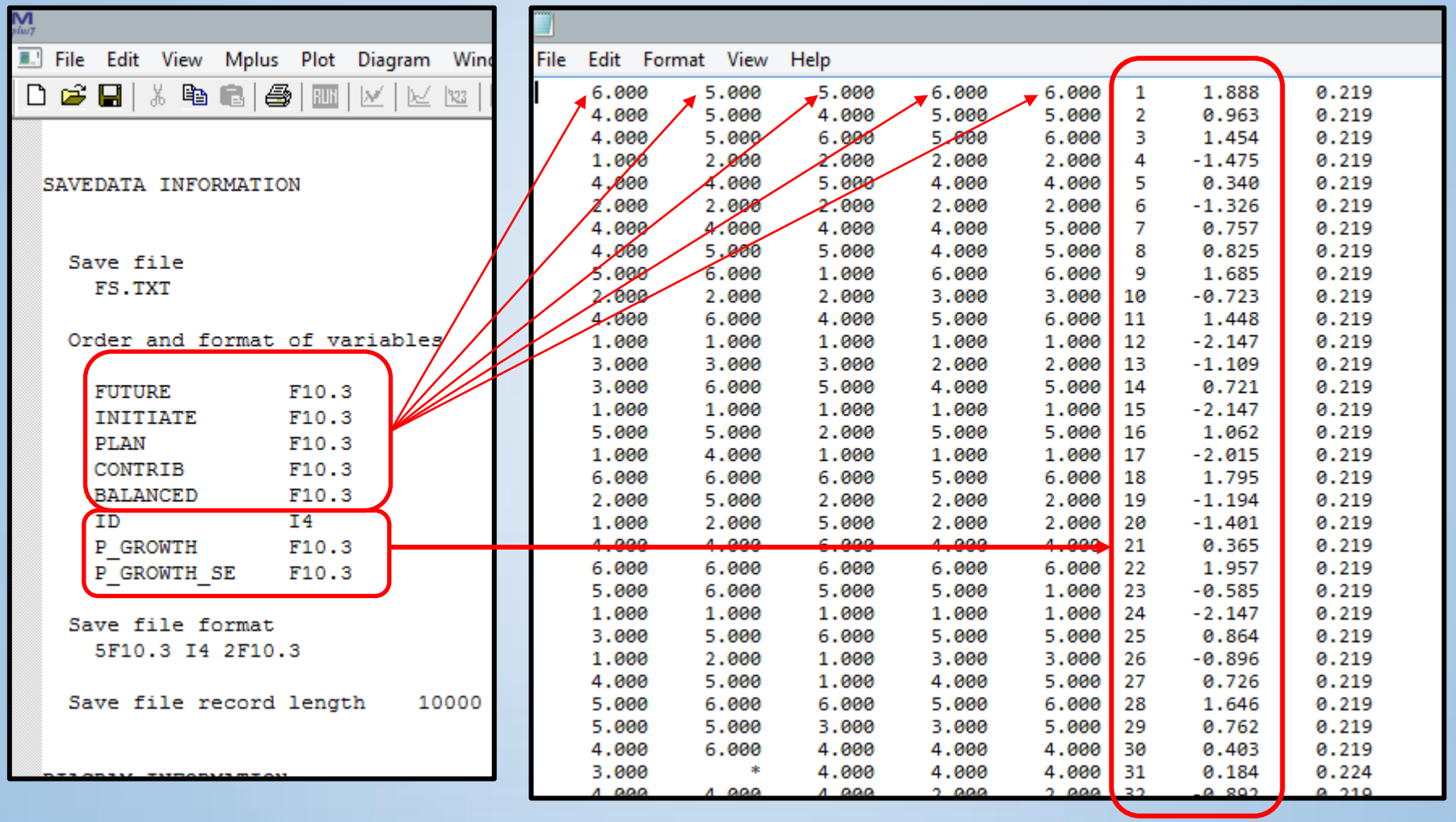
The data



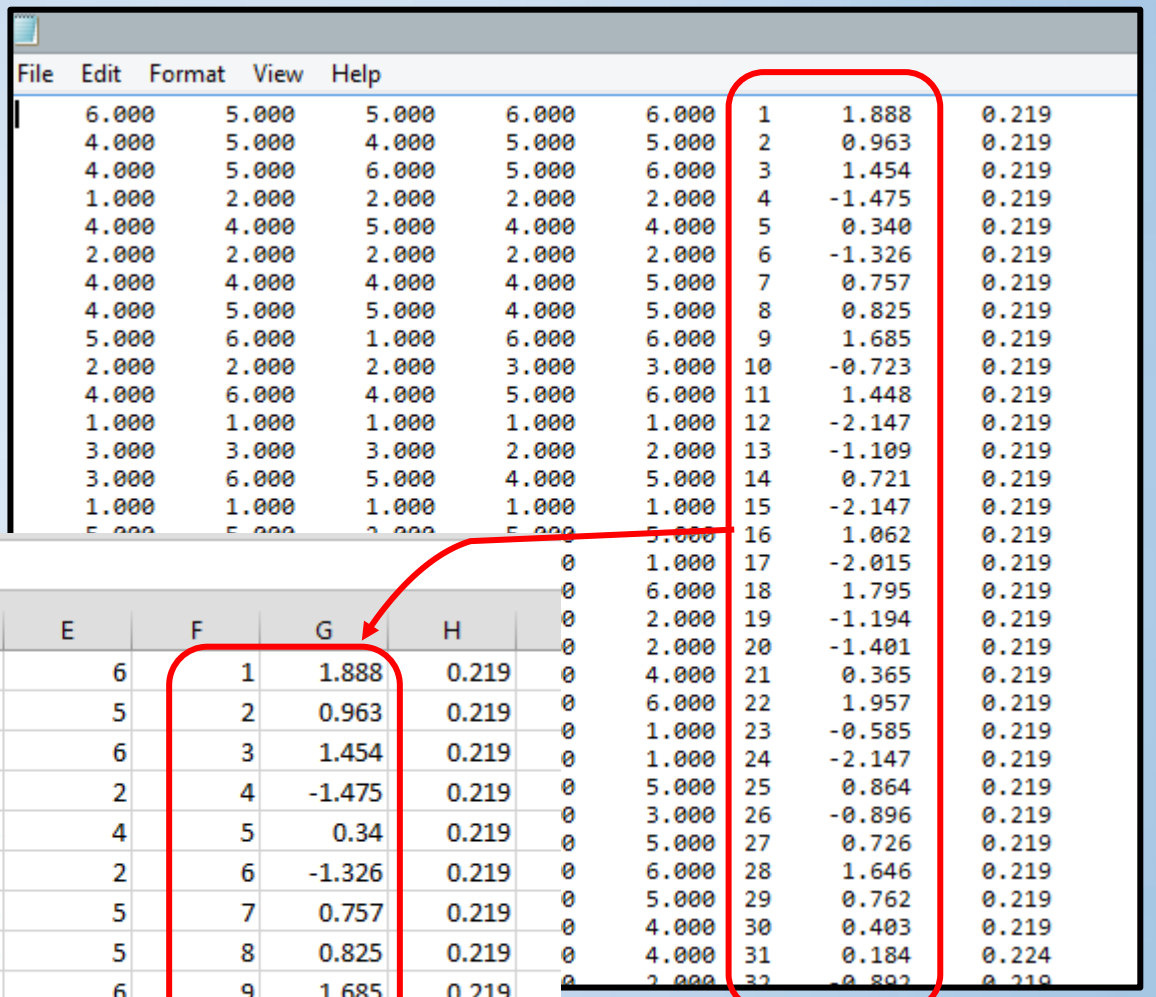
	ID	sense_future	initiate_transition	action_plan	unique_contribution	balanced_life
1	1	1	1	3	1	1
2	2	4	2	2	2	2
3	3	3	3	3	3	2
4	4	4	3	4	4	4
5	5	5	5	5	2	6
6	6	6	6	6	6	5
7	7	1	1	3	1	1
8	8	4	2	2	2	2
9	9	3	3	3	3	2
10	10	4	3	4	4	4
11	11	5	5	5	2	6
12	12	6	6	6	6	5

Let’s do it together…

Factor scores



Factor scores



A1 | 6

	A	B	C	D	E	F	G	H
1	6	5	5	6	6	1	1.888	0.219
2	4	5	4	5	5	2	0.963	0.219
3	4	5	6	5	6	3	1.454	0.219
4	1	2	2	2	2	4	-1.475	0.219
5	4	4	5	4	4	5	0.34	0.219
6	2	2	2	2	2	6	-1.326	0.219
7	4	4	4	4	5	7	0.757	0.219
8	4	5	5	4	5	8	0.825	0.219
9	5	6	1	6	6	9	1.685	0.219
10	2	2	2	3	3	10	-0.723	0.219
11	4	6	4	5	6	11	1.448	0.219
12	1	1	1	1	1	12	-2.147	0.219

Factor scores

	id	p_growth
1	1.00	1.89
2	2.00	.96
3	3.00	1.45
4	4.00	-1.48
5	5.00	.34
6	6.00	-1.33
7	7.00	.76
8	8.00	.83
9	9.00	1.69
10	10.00	-.72
11	11.00	1.45
12	12.00	-2.15
13	13.00	-1.11

A1 | 6

	A	B	C	D	E	F	G	H
1	6	5	5	6	6	1	1.888	0.219
2	4	5	4	5	5	2	0.963	0.219
3	4	5	6	5	6	3	1.454	0.219
4	1	2	2	2	2	4	-1.475	0.219
5	4	4	5	4	4	5	0.34	0.219
6	2	2	2	2	2	6	-1.326	0.219
7	4	4	4	4	5	7	0.757	0.219
8	4	5	5	4	5	8	0.825	0.219
9	5	6	1	6	6	9	1.685	0.219
10	2	2	2	3	3	10	-0.723	0.219
11	4	6	4	5	6	11	1.448	0.219
12	1	1	1	1	1	12	-2.147	0.219