All softwares are wrong, but some are useful

Francisco Cardozo
University of Miami

Eric Brown
University of Miami

February 9, 2024

Agenda

- R and *M*plus
- Data set
- Multitrait-Multimethod Matrix
- Pre-processing
- CFA

R and *M*plus

R and *M*plus

R

- Integrate your data, analysis and text
- Tools for data manipulation
- Packages
- Free

***M*plus**

- Consistent syntax
- Informative outputs
- Specialized for SEM models
- Awesome Blog

Integrate both

You can use R to generate:

- Data sets to *M*plus format
- Create *M*plus code
- Run *M*plus code
- Read *M*plus outputs

Keep your work on *M*plus and use R to write the report.

> *M*plus -> R

Before jumping to *M* plus...

*M*plus requires that your data is almost ready for analysis. You can do pre-processing in *M*plus (like filter of computing variables) but it is not the best tool for that work.

R is great for data pre-processing.

> R -> *M*plus -> R

R has packages to do SEM Analysis

- lavaan
  - CFA
  - Growth Curves
  - Mediation
  - Multilevel SEM
- polca
  - Latent Class Analysis
- simsem
  - Simulated Structural Equation Modeling

Example

Prevention of alcohol use

We usually measure risk factors by asking students in the schools, but because of Cov

We wanted to ask while they were at home, but it was extremely complicated becaus

> *What happens if we ask parents?* 🤔

Is the parent evaluation of the student's risk a good approximation of the student's ris

What is the correlation between students' and parents' perceptions of risk?

Data set

The data set contains:

- Risk and protective factors
- Parent's and students' perspective

Community Disorganization (CRCDO)

How much do each of the following statements describe your neighborhood?

Var_name	Child	Parent	4	3	2	1
NHDELIN	Crime	Crime	YES!	Yes	No	NO!
NHVENTA	Drugs	Drugs	YES!	Yes	No	NO!
NHFIGHT	Fights	Fights	YES!	Yes	No	NO!
NHEMPTY	Lots of empty or abandoned buildings	Lots of empty or abandoned buildings	YES!	Yes	No	NO!
NHGRAF	Lots of graffiti	Lots of graffiti	YES!	Yes	No	NO!

Perceived Risks of Drug Use (PRPRD)

How much do you think people risk harming themselves (physically or in other ways) if they…

Var_name	Child	Parent	4	3	2	1
HMCIG	…smoke one or more packs of cigarettes per day?	…smoke one or more packs of cigarettes per day?	None	Small	Moderate	Big
HMMARO	…try marijuana once or twice?	…try marijuana once or twice?	None	Small	Moderate	Big
HMMARR	…smoke marijuana regularly?	…smoke marijuana regularly?	None	Small	Moderate	Big
HMALC	…take one or two drinks of an alcoholic beverage (beer, wine, or liquor) nearly every day?	…take one or two drinks of an alcoholic beverage (beer, wine, or liquor) nearly every day?	None	Small	Moderate	Big

Poor Family Management (FRPFM)

In your family…

Var_name	Child	Parent	1	2	3	4
FAMRULE	The rules in my family are clear.	The rules in my family are clear.	YES!	Yes	No	NO!
HMWORK	My parents ask if I've gotten my homework done.	I frequently ask my children if they have finished their homework.	YES!	Yes	No	NO!
PARKNOW	When I am not at home, one of my parents knows where I am and who I am with.	When my children are not at home, one of the adults knows where they are and who they are with.	YES!	Yes	No	NO!
CMHOME	Would your parents know if you did not come home on time?	I would realize if my children do not arrive on time at home.	YES!	Yes	No	NO!
CLRRULE	My family has clear rules about alcohol and drug use.	I would realize if my children do not arrive on time at home.	YES!	Yes	No	NO!
CATCHAL	If you drank some beer or wine or hard liquor (for example, vodka, whiskey or gin) without your parent's permission, would you be caught by your parents?	If my son/daughter drinks beer, brandy, rum, wine, cocktails, or any other liquor without permission, I would realize it	YES!	Yes	No	NO!
CATCHSK	If you skipped school, would you be caught by your parents?	If my children skip school, I would realize it.	YES!	Yes	No	NO!

Multitrait-Multimethod Matrix

Multimethod= Parent and Child

Multitrait = Community Disorganization, Poor Family Management, and Perceived Risks of Drug Use

Cronbach's Alpha

Community Disorganization = 0.75

Perceived Risk of Drug Use = 0.662

Poor Family Management = 0.818

Cronbach's Alpha Community Disorganization (CRCDO)

Parent, student

```
[1] 0.7383984
```

```
[1] 0.7579803
```

Cronbach's Alpha Perceived Risks of Drug Use (PRPRD)

Parent, student

```
[1] 0.6401554
```

```
[1] 0.618836
```

Cronbach's Alpha Poor Family Management (FRPFM)

Parent, student

```
[1] 0.8343457
```

```
[1] 0.8081604
```

MultiTrait MultiMethod Matrix

	P_CRCDO	P_FRPFM	P_PRPRD	S_CRCDO	S_FRPFM	S_PRPRD
P_CRCDO		Corr: -0.126.	Corr: -0.032	Corr: 0.415***	Corr: -0.089	Corr: -0.074
P_FRPFM			Corr: -0.085	Corr: -0.074	Corr: 0.161*	Corr: -0.012
P_PRPRD				Corr: -0.078	Corr: 0.010	Corr: 0.267***
S_CRCDO					Corr: 0.020	Corr: 0.088
S_FRPFM						Corr: -0.144*
S_PRPRD						

CRCDO= Community Disorganization, **PRPRD**= Perceived Risks of Drug Use, **FRPFM**= Poor Family Management

Pre-processing

Data set in wide format

You have an ID column, but the participants (unit of analysis) only appear once.

ID	Parent_RF1_item1	Parent_RF1_item2	Student_RF1_Item1	Student_RF1_Item2
1	1	3	2	1
2	4	2	3	1

Data set in long format

You have ID column, and the participants are repeated multiple times in different rows.

ID	Member	Risk Factor	Item	Answer	Label
1	Parent	RF1	Item1	2	no
1	Parent	RF1	Item2	3	yes
1	Student	RF1	Item1	2	no
1	Student	RF1	Item2	1	YES!
1	Parent	RF2	Item1	4	NO!
1	Parent	RF2	Item2	2	no
1	Student	RF2	Item1	3	yes
1	Student	RF2	Item2	2	no
2	Parent	RF1	Item1	4	NO!
2	Parent	RF1	Item2	2	yes
2	Student	RF1	Item1	3	no
2	Student	RF1	Item2	1	YES!

Long and wide format

Click to go to the Practice

MplusAutomation

MplusAutomation

A package:

- Run *M*plus scripts from R.
- Convert data into *M*plus format.
- Generate *M*plus scripts
- Read *M*plus outputs into R
- Plot model results

MplusAutomation

```r
# 1. Install to save the package in your computer.
# You only need to do this once
install.packages("MplusAutomation")

# 2. Call the package to activate the functions saved in the packag
library(MplusAutomation)
```

MplusAutomation

Before you start pivoting your data, begin by…

Import your data

- bulkreadr
- haven
- readxl

```
1 library(readxl) # To read excel files in R
2 library(tidyverse) # To activate the "pivot" functions
3 library(MplusAutomation) # To activate the function "prepareMplusDa
4
5 dyads <- read_xlsx("./replace_this_with_the_name_of_your_data_file.
```

MplusAutomation

Reshape data from long to wide format.

```r
# Pivot wider your data
Mplus_dyads <-
  dyads |>
  filter(FACTORNAME %in% c("CRCDO", "PRPRD", "FRPFM")) |>
  select(ID, MEMBER, VARNAME, answer) |>
  pivot_wider(names_from = c(MEMBER, VARNAME),
              values_from = answer)

Mplus_dyads |> head(2) |> gt::gt()
```

ID	P_SAFENH	P_NHDELIN	P_NHVENTA	P_NHFIGHT	P_NHEMPTY	P_NHGRAF	P_FAMRULE	P_HMWORK	P_PARKNOW	P_CMHOM
01-606-001-20102021	1	2	1	2	2	2	4	4	3	
01-606-002-20102021	3	2	2	2	3	3	3	4	4	

MplusAutomation

Recode missing data

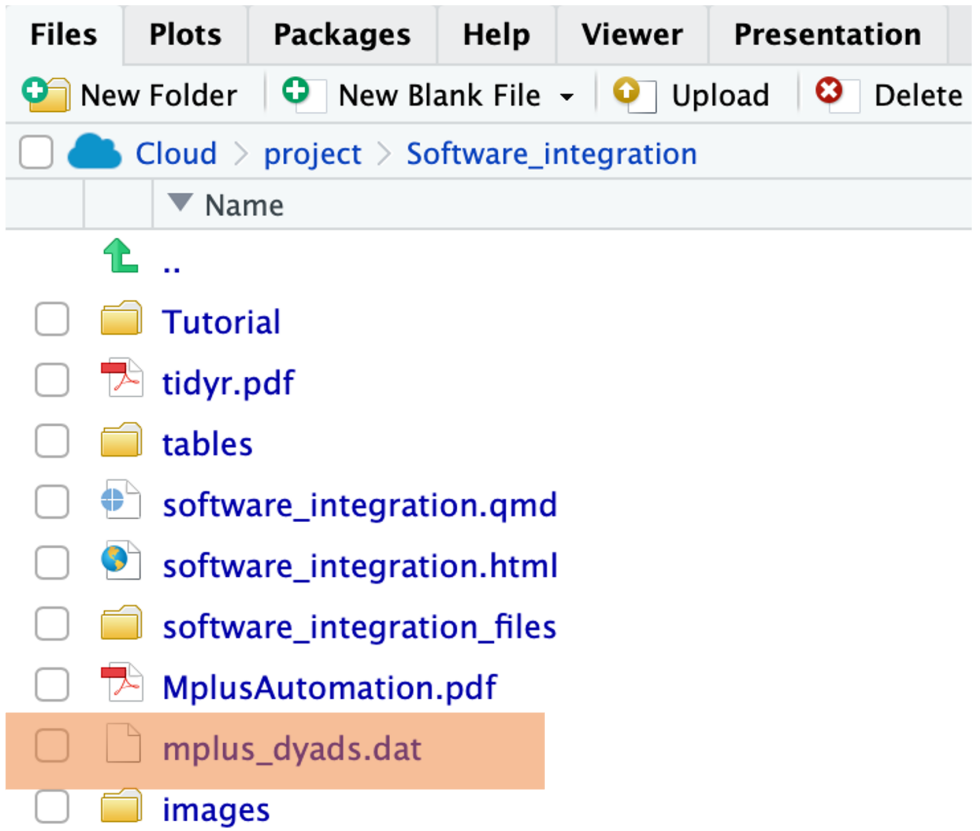
```
# Define missing data value
Mplus_dyads[is.na(Mplus_dyads)] <- 999
```

And export

```
# Export your data to Mplus
prepareMplusData(Mplus_dyads, "Mplus_dyads.dat")
```

MplusAutomation

```
> # Define missing data value
> mplus_dyads[is.na(mplus_dyads)] <- 999
> # Export your data to Mplus
> prepareMplusData(mplus_dyads, "mplus_dyads.dat")
TITLE: Your title goes here
DATA: FILE = "mplus_dyads.dat";
VARIABLE:
NAMES = ID Parent_CRCDO Parent_FRPFM Parent_PRPRD Student_CRCDO Student_FRPFM
     Student_PRPRD;
MISSING=.;
```



Practice

CFA in R using Lavaan

CFA Lavaan

To estimate the model you need to

1. use `library(lavaan)`
2. Specify your model

Perceived Risks of Drug Use (PRPRD) by items HMALC, HMCIG, HMMARO and HMMARR.

```
1 library(lavaan)
2
3 model  <- '
4
5 PRPRD =~ HMALC  + HMCIG + HMMARO + HMMARR
6
7 '
```

CFA Lavaan

To estimate the model you need to

3. Estimate your model

```
1 # Entire sample
2 cfa_fit <- cfa(model, data=Mplus_cfa)
3
4 cfa_results <- summary(cfa_fit, standardized=TRUE, fit.measures=TRU
```

The standardized argument equal TRUE gives you results equivalent to STD (Std.lv) and STDYX (Std.all) on Mplus.

You can have reduced output using `standardizedsolution()`

lavaan ouput

```
lavaan 0.6.16 ended normally after 30 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         8

                                                  Used       Total
  Number of observations                           386         388

Model Test User Model:

  Test statistic                                45.153
  Degrees of freedom                                 2
  P-value (Chi-square)                           0.000
```

lavaan ouput

	lhs	op	rhs	est.std	se	z	pvalue	ci.lower	ci.upper
1	PRPRD	=~	HMALC	0.448	0.049	9.222	0.000	0.353	0.543
2	PRPRD	=~	HMCIG	0.417	0.049	8.480	0.000	0.321	0.513
3	PRPRD	=~	HMMARO	0.562	0.047	12.009	0.000	0.470	0.654
4	PRPRD	=~	HMMARR	0.926	0.051	18.296	0.000	0.827	1.025
5	HMALC	~~	HMALC	0.800	0.043	18.399	0.000	0.714	0.885
6	HMCIG	~~	HMCIG	0.826	0.041	20.157	0.000	0.746	0.907
7	HMMARO	~~	HMMARO	0.684	0.053	13.013	0.000	0.581	0.787
8	HMMARR	~~	HMMARR	0.143	0.094	1.523	0.128	-0.041	0.326
9	PRPRD	~~	PRPRD	1.000	0.000	NA	NA	1.000	1.000

My First CFA in R

Advanced Research Methods

LHS	OP	RHS	EST.STD	SE	Z	PVALUE	CI.LOWER	CI.UPPER
PRPRD	=~	HMALC	0.448	0.049	9.222	0.000	0.353	0.543
PRPRD	=~	HMCIG	0.417	0.049	8.480	0.000	0.321	0.513
PRPRD	=~	HMMARO	0.562	0.047	12.009	0.000	0.470	0.654
PRPRD	=~	HMMARR	0.926	0.051	18.296	0.000	0.827	1.025
HMALC	~~	HMALC	0.800	0.043	18.399	0.000	0.714	0.885
HMCIG	~~	HMCIG	0.826	0.041	20.157	0.000	0.746	0.907
HMMARO	~~	HMMARO	0.684	0.053	13.013	0.000	0.581	0.787
HMMARR	~~	HMMARR	0.143	0.094	1.523	0.128	−0.041	0.326
PRPRD	~~	PRPRD	1.000	0.000	NA	NA	1.000	1.000

Practice

CFA by groups

Parents

```r
parents <-dyads |> filter(MEMBER=="P") |>
  select(ID, answer, VARNAME) |>
  filter(VARNAME %in% c("HMALC", "HMCIG", "HMMARO", "HMMARR")) |>
  pivot_wider(names_from = VARNAME, values_from=answer, id_cols = I
  select(-ID)

parents_results <- cfa(model, data=parents)  |>
  summary(standardized=TRUE, fit.measures=TRUE)
```

Parents

```
lavaan 0.6.16 ended normally after 27 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         8

  Number of observations                           194

Model Test User Model:

  Test statistic                                24.648
  Degrees of freedom                                 2
  P-value (Chi-square)                           0.000

Model Test Baseline Model:
```

Students

```r
students <-dyads |> filter(MEMBER=="S") |>
  select(ID, answer, VARNAME) |>
  filter(VARNAME %in% c("HMALC", "HMCIG", "HMMARO", "HMMARR")) |>
  pivot_wider(names_from = VARNAME, values_from=answer, id_cols = I
  select(-ID)

students_results <- cfa(model, data=students)  |>
  summary(standardized=TRUE, fit.measures=TRUE)
```

Students

```
lavaan 0.6.16 ended normally after 34 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         8

                                                  Used       Total
  Number of observations                           192         194

Model Test User Model:

  Test statistic                                26.591
  Degrees of freedom                                 2
  P-value (Chi-square)                           0.000
```

We can do better

Are we measuring the same thing? Invariance? Dyads?

The end