Intro

Introduction I

Today you'll be working together as a group to better understand multivariate twin models, including the Independent Pathway Model (IPM) and Common Pathway Model (CPM) and how these are estimated using OpenMx.

INSTRUCTIONS:

1) We will use a workbook (Word file) as the basis of this tutorial, but you are asked to answer the questions that are in the workbook in this survey; this allows the tutors to monitor your progress. Make sure that all of you have the workbook open during the tutorial.

2) Only one person should fill in the survey. It may work best if another person shares their screen to run the chunks of R script for the whole group to see.

3) Throughout the practical, you will be asked to work out solutions to problems and to share your solutions using this survey. Please discuss and debate your answers among yourselves before asking the Scribe to put down your group answer in Qualtrics. Everyone should try to participate. After you have collectively come up with your best answer, put that answer in the answer box.

4) The workbook asks you to copy and paste pieces of R script. Everyone should have R Studio open on their computer and everyone should run it together at the same time. When prompted in the workbook, run the relevant Section of the R script line by line (e.g., hitting Ctr + Enter, or Apple + Enter) or in small chunks of code. The R scripts are broken up into Sections. Please don't skip ahead - keep pace with the Qualtrics survey. Everyone should work through the survey and script together.

5) No cheating: Don't skip ahead in the survey or in the workbook to get the answers. The point of all this isn't to get a good "score" - it's to learn! **It is important to stop reading in the workbook when you have reached a question, as sometimes the**

answer to the question is provided shortly after the question to keep things going.

6) Relatedly, it is easy to just cut and paste the R code and keep going, but try to understand what the code says, what is actually being calculated and whether you understand the syntax, i.e. what you are doing!

7) Try to get through all sections, but there is no need to rush. If you are unable to finish the whole practical in the allotted time you can finish it later by yourself.

8) After Session B, a workbook with answers will be provided in the folder of today.

Introduction II

Please spend a couple of minutes introducing yourselves. You will need to then choose two people who will serve two different roles in your group:

As you're going through this tutorial, we'll call you back into the main room once or twice to check in and take a break.

Answers to some questions come later on in the workbook or the R script (remember, no cheating). If you need more help please click the 'Ask for help' button in Zoom and someone will come.

Please appoint someone to be filling in this survey now.

FIRST QUESTION (to be answered by this ONE person, in your room): Which zoom breakout room are you in (what number)?

[   ]

What are the first names of your group members?

[   ]

Where are you from?

[   ]

Has anyone in your group run a multivariate (i.e., >1 variable/trait) twin model, independent pathway model and/or common pathway model before? (multiple answers are possible)

☐ Group members have run a multivariate twin model

☐ Group members have run an independent pathway model

☐ Group members have run a common pathway model

☐ Group members have not run either of these types of analyses before

Getting started

If you haven't already done so, open the workshop computing environment
https://workshop.colorado.edu

Open the workshop SSH client
Create a directory to hold today's work
mkdir day5

Change into that directory, and then copy over the exercises.
cd day5
cp /faculty/dirk/day5/* .

# Open the workshop Rstudio client
# Open the folder day5 (bottom right quadrant of the screen)
# Set this folder as your working directory, either with the command
setwd('~/day5')
# Or by going to Session > Set Working Directory > Choose Directory
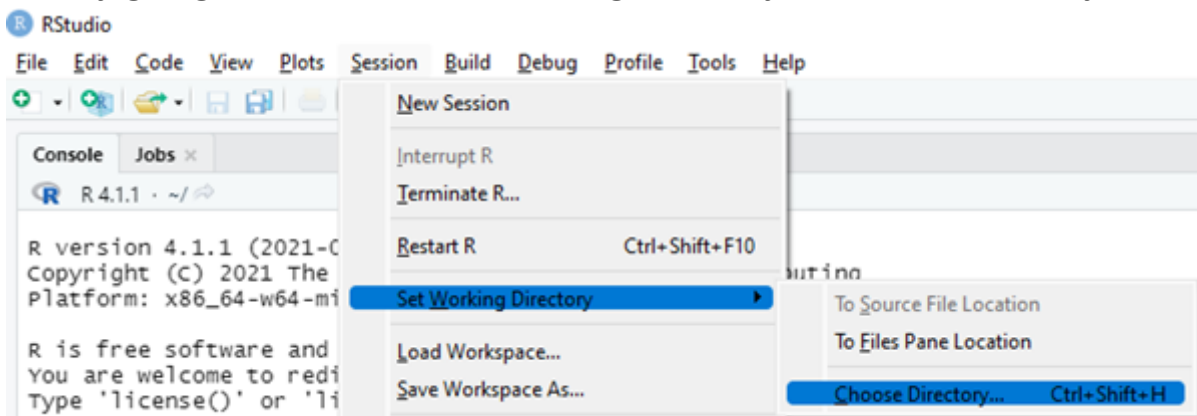
Open a script window in R studio
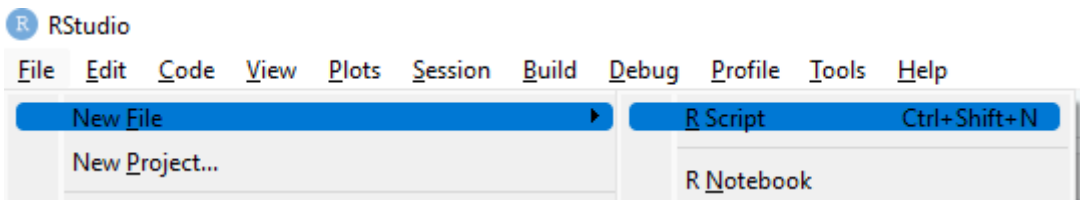
Copy and paste the R code from the Word workbook to the R script window, select and execute by clicking on →Run in the top right corner of the R script window.
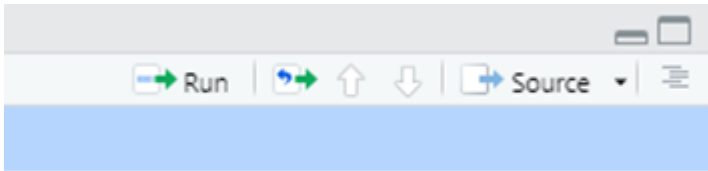
Workbook

You can find the workbook following the link below (please right click and open in new tab or window):

Workbook

If you completed the previous steps, you should have also copied the workbook to your directory (workbook_Multivariate twin models_v2_students.docx), which you can access after the practical. For those of you who are unable to open Word files, there is also a PDF file in the folder.

The R code that is to be cut-and-pasted is shown between #-> and #<- in the Word workbook.

In the Word workbook, all text rendered in red font `Courier New` is R code input, e.g.,
```
#
print(c("Hello World"))
#
```

In this document, all text rendered in blue font `Courier New` is R code output, e.g.,

```
> print(c("Hallo World"))
[1] "Hallo World
```

Practical - Part I

Skinfold data. The 4-variate ADE twin model. 15 questions

Let's get started with the practical that accompanies the 5 PPT presentations on multivariate twin modelling, based on the Independent Pathway Model (IPM) and the Common Pathway Model (CPM). In this practical, we will analyze the following two data sets:

Data set	Content
skintwindata	skinfold data, skinfold measured at 4 locations
ntwindata	Item scores of 4 (5 point scale) items designed to measure neuroticism (a personality trait)

The *ntwindata* and the *skintwindata* featured in the 5 PPTs. The present aim is to reproduce the results in the PPTs ourselves using the OpenMx library in R Studio. In the first part of this practical, we analyze the *skintwindata*. Before the practical, we briefly go over multivariate modelling and the IPM and CPM.

Here is an overview of what we hope to do in part 1 of this practical

- Read the skinfold data (skintwindata) into R studio, and obtain basic descriptives in the MZ and the DZ samples
- Calculate the MZ and DZ twin correlations and use Falconer's equations to obtain a rough guess of ACE or ADE standardized variance components
- Consider the relationship between the covariance matrix and the correlation matrix

In the PPT presentation, we first fitted a bivariate model of skinfold measured at the biceps and triceps. Here we will proceed straight away to the 4 variate model.

- Calculate the MZ and DZ covariance matrices and means
- Calculate the MZ and DZ covariance matrices in OpenMx with sex as a covariate
- Fit the 4 variate ADE twin model and consider the results

Multivariate ADE model

Question 1.1.

Check out the MZ and DZ phenotypic correlation (rmz, rdz) of each phenotype. The rule of thumb is: if 2*rdz>rmz, then choose an ACE model; if 2*rdz<rmz choose a ADE model. What are the correlations?

Fill in the correlation values for the 4 phenotypes below. Make sure to save them in an object or write them down because you need these values later on.

	MZ	DZ
bic		
tri		
ssc		
sil		

Based on these correlations, what model would you choose?

Question 1.2.

Using "Falconer's equations" we can obtain preliminary estimates of the standardized A, D, E variance components, VA, VD, & VE. In the case of the ADE model, these equations are:

VA = 4*rdz-rmz
VD= 2*rmz-4*rdz
VE= 1-VA-VD

Apply these to the correlations, to obtain the standardized components VA and VD. How much of the phenotypic variance is due to A (additive genetic effects), D (non-additive genetic effects) and E (non-shared environmental effects)?

	A (in %)	D (in %)	E (in %)
bic			
tri			
ssc			
sil			

Question 1.3.

In the output, we can see the 8x8 covariance matrix of the MZ group (called Smz4) and

the correlation matrix of the MZ group (called Rmz4). We can express the covariance matrix as a function of the correlation matrix Rmz4 and the diagonal matrix containing the standard deviations. Do this in R as follows 1) get the standard deviations from the covariance matrix, and put these in the diagonal of a 8x8 matrix, called Sdmz and carry out the multiplication: Sdmz%*%Rmz4%*%t(Sdmz)

Verify that the covariance and correlations matrices that you calculate in different ways are exactly the same.

O Done!

Question 1.4.

Repeat the analyses in the twin 2 member. Check out the Multiple R-squared statistic and the p value associated with the regression coefficient in the regression of the skinfold measures on sex (Pr(>|t|)).

What do you conclude with respect to the main effect of sex?

[   ]

Question 1.5.

What are the MZ and DZ correlations of the 4 phenotypes based on the OpenMx output?

	MZ	DZ
bic	[ ]	[ ]
tri	[ ]	[ ]
ssc	[ ]	[ ]
sil	[ ]	[ ]

Why do they differ from those of question 1.2?

[   ]

Question 1.6.

What do you conclude on the basis of the MZ and DZ cross-trait – cross-twin

correlations of biceps and triceps (underlined), given MZ cor(bic_T1, tri_T2) > given DZ cor(bic_T1, tri_T2)?

MZ cross-twin-cross-trait correlations		DZ cross-twin-cross-trait correlations	
	bic_T1 tri_T1		bic_T1 tri_T1
bic_T2	**0.769** 0.706	bic_T2	**0.312** 0.269
tri_T2	0.737 **0.795**	tri_T2	0.254 **0.270**

Question 1.7.

What does the guess svVA=svS*.4 actually imply with respect to the variance of the 4 phenotype phenotypes? Is this reasonable given your answer to question 1.2.? (From question 1.2., we know that the percentages are 50.9%, 28.0%, 37.7% and 32.9%).

Question 1.8.

Verify that VA_ = sA_%*%RA_%*%sA_.

○ Done!

Question 1.9.

Extract the covariance matrices of A (already done!), D and E, call these VA_, VD_, and VE_. But here we only need the 2x2 matrices of the phenotypes biceps and triceps.

○ Done!

Question 1.10.

Look at the values in VA_, VD_, and VE_, they should resemble approximately the values in *Figure slide 26 left*. NOTE: they are not exactly equal, because the results in the slides were obtain by fitting the bivariate model. but the present results were obtained by fitting the 4-variate model. **Verify that the parameters in *Figure slide 26 left* resemble the values in VA_, VD_, and VE_.**

○ Done!

Question 1.11.

Verify that the parameter in *Figure slide 26 right* resemble your values.

○ Done!

Question 1.12.

Verify that your parameter values resemble those in Figure slide 27 right.

○ Done!

Question 1.13.

If I tell you that in the model the environmental correlation equals .789, does that mean that E is contributing greatly to the phenotypic correlation? The actual expected phenotypic correlation is .893. Answer this by referring to the actual contribution of E to the phenotypic correlation.

```

```

Question 1.14.

What is "wrong" with the 4x4 correlation matrix of D?

```

```

Question 1.15

Use the omxSetParameters() function to fix the D correlations to 1.0, and use the mxCompare() function to test whether these constraints are ok statistically. Write down the values **of the comparison** below:

Values

difLL

Values

diffdf

p-value

If OK, what does this result suggest (interpretation)?

Why is df=6 (diffdf=6)?

IPM and CPM

Practical Part II

Four neuroticism Items. Multivariate ADE model. The AE Independent Pathway model and the Common Pathway Model. 12 questions.

In the second part of the practical, we will analyze 4 neuroticism items in the following steps
1) Read the data into R studio and create the MZ and DZ data frames
2) Calculate descriptives and correlations
3) Fit the 4 variate ADE model, test the significance of the D component
4) Fit the AE IPM, interpret the results
5) Fit the AE CPM, interpret the results
6) Fit the AE model to the Neuroticism sum score, compare results with those of step 5

The data file, which is called *ntwindata*, contains the neuroticism items scores in a sample of 1528 MZ pairs and 1277 DZ pairs. The response format of the items is a 5-point scale (1 to 5), the higher the item score, the more neurotic. The dataset includes the variable zyg (zygosity, zyg = 1 for MZ; zyg = 2 for DZ).

Question 2.1

Consider the twin correlations of each item. Remember that rmz>2*rdz suggests an ADE model and rmz<2*rdz suggests an ACE model. However, rmz=2*rdz suggests an

AE model and rmz = rdz suggests a CE model. Based on the correlations, which model is most likely to hold?

```
[                                                                            ]
```

Question 2.2.

Fill in the values of the likelihood ratio test below.

	Value
diffLL	
diffdf	
p-value	

Based on these values: are the means equal in the MZs and DZs?

```
[                                                                            ]
```

And why is df=4 (diffdf=4)?

```
[                                                                            ]
```

Question 2.3.

Fill in the values of the LRT below to answer the following question.

	Value
diffLL	
diffdf	
p-value	

Does the ADE model fit relative to the saturated model?

```
[                                                                            ]
```

Question 2.4.

Use omxSetParameters() to drop the D component from the ADE model, and fit the AE 4 variate model. The D component parameter labels are given in the OpenMx script.

○ Done!

Question 2.5.

Use mxCompare() to carry out the Likelihood ratio test that the D component is zero (fitADE4V vs fitAE4V).

	Value
diffLL	☐
diffdf	☐
p-value	☐

What do you conclude based on the output?

☐

And why is df=10 (diffdf=10)?

☐

Question 2.6.

You can see that the A correlations are much higher (.713 to .932) than the E correlations (.263 to .349). Does that mean that A contributes more to the phenotypic correlations that does E? Check your answer against the output of the following code. First, calculate the covariance matrices SA and SE (yes you actually already have these!). Second, get the proportions, and answer the question in the light of these proportions.

☐

Question 2.7.

Does the AE IPM model fit the data relative to the 4 variate AE model?

	Value
diffLL	☐
diffdf	☐
p-value	☐

Why is df=4 (diffdf = 4)?

[ ]

Question 2.8.

Consider the A and E item residual variances. **With respect to item specific A and E effect, what do you conclude?**
Bear in mind that the E residual variance include measurement error. For instance in the practical part 1, we analysed skinfold data. The correlation between skinfold measures and ultrasound measures of subcutaneous fat is between .80 and .90 (the ultrasound measurement is the golden standard). That means that the skinfold measures are not perfect measures of subcutaneous fat, i.e., they are characterized by measurement error. The same applies to the neuroticism items.

[ ]

Question 2.9.

The AE IPM has 4 means, 2x4 (A and E) factor loadings and 2x4 (A and E) residual variances, i.e., 20 parameters. How many independent parameters does the CPM model have?

[ ]

Why is the df in mxCompare(fitAE4IPM,fitAE4CPM2) equal to 3?

[ ]

Question 2.10.

Does the AE CPM (version B) fit the data relative to the AE IPM?

	Value
diffLL	[ ]
diffdf	[ ]
p-value	[ ]

Question 2.11.

What is the heritability of the latent variable Neuroticism, according to the fitAE4CPM2 output?

```
```

Question 2.12.

What is the heritability of the Neuroticism sum score?

```
```

Why does it differ from the heritability based on the CPM (see answer to question 2.11)?

```
```

End

Congratulations, you have successfully completed this practical!

We hope you expanded your model fitting expertise in OpenMx and learned how to test multivariate twin models, the independent pathway model and common pathway model.

If you have questions on today's practical, feel free to post them on the forum.

Powered by Qualtrics