Session 2: SEM | Study Design | Phenomics - Practical 2

twin Modeling in OpenMx from expected to actual relatedness

1. Create day2/practical2 directory
2. Copy all files from faculty/hmaes/2021/day2/practical2 to your directory
3. Determine your level of OpenMx proficiency:
   Beginner: first time using OpenMx
   Intermediate: attended OpenMx workshop or used existing OpenMx scripts
   Advanced: written OpenMx scripts or adapted existing ones extensively
4. Note that filenames are color coded in orange below.
5. Objects from model 1 that are not changed in models 2-6 are re-used, but do not need to be repeated, as long as they are included in the mxModel statement(s).

For Beginners

1. Open oneACEvc7.R or open individual scripts 1: oneACEvc_1cov.R
2. Make sure you have data: tsDataS7.txt and miFunctions.R in the same directory

3. Run **Model 1** (all lines up 111)
4. Inspect the lines related to specifying the expected covariance matrices

```
# Create Matrices for Variance Components
covA       <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svVa,
 label="VA11", name="VA" )
covC       <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svVa,
 label="VC11", name="VC" )
covE       <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svVe,
 label="VE11", name="VE" )

# Create Algebra for expected Variance/Covariance Matrices in MZ & DZ twins
covP       <- mxAlgebra( expression= VA+VC+VE, name="V" )
covMZ      <- mxAlgebra( expression= VA+VC, name="cMZ" )
covDZ      <- mxAlgebra( expression= 0.5%x%VA+ VC, name="cDZ" )
expCovMZ   <- mxAlgebra( expression= rbind( cbind(V,       cMZ),
                                            cbind(t(cMZ), V)), name="expCovMZ" )
expCovDZ   <- mxAlgebra( expression= rbind( cbind(V,       cDZ),
                                            cbind(t(cDZ), V)), name="expCovDZ" )
```

5. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals. You can find these in the regular summary of the fitted model object: summary(fitACEvc)

```
> summary(fitACE1)
Summary of oneACEvc_1cov

free parameters:
    name    matrix row col     Estimate    Std.Error A
1 interC intercept   1   1 0.0745740554 0.0654497596
2  betaS        bS   1   1 0.0981537390 0.0308553226
3  betaA        bA   1   1 0.0016269288 0.0020417776
4   VA11        VA   1   1 0.4128498423 0.0480040225
5   VC11        VC   1   1 0.1907156678 0.0443951969
6   VE11        VE   1   1 0.2762940993 0.0122050017
```

```
confidence intervals:
                          lbound   estimate     ubound note
oneACEvc_1cov.US[1,1] 0.32104727 0.41284984 0.51000018
oneACEvc_1cov.US[1,2] 0.10186692 0.19071567 0.27644456
oneACEvc_1cov.US[1,3] 0.25366495 0.27629410 0.30163541
oneACEvc_1cov.US[1,4] 0.36561684 0.46922241 0.57740652
oneACEvc_1cov.US[1,5] 0.11689735 0.21675693 0.30964261
oneACEvc_1cov.US[1,6] 0.28575636 0.31402066 0.34519481

Model Statistics:
               |  Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
       Model:              6                  3994              9975.7666
   Saturated:             NA                    NA                     NA
Independence:             NA                    NA                     NA
Number of observations/statistics: 2000/4000

Information Criteria:
      |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
AIC:       1987.7666              9987.7666               9987.8088
BIC:     -20382.2378             10021.3720              10002.3097
CFI: NA
TLI: 1   (also known as NNFI)
RMSEA:  0  [95% CI (NA, NA)]
Prob(RMSEA <= 0.05): NA
To get additional fit indices, see help(mxRefModels)
timestamp: 2021-06-04 09:57:54
Wall clock time: 27.219256 secs
optimizer:  NPSOL
OpenMx version number: 2.19.5.1
Need help?  See help(mxSummary)
```

6. or get them using one of the helper functions: fitGofs (for goodness-of-fit statistics) and fitEstCIs (for unstandardized variance components VA, VC & VE , standardized variance components SA, SC, SE (in mxMatrix US)).

```
> fitGofs(fitACE1)
Mx:oneACEvc_1cov  os=4000  ns=2000   ep=6   co=0  df=3994  ll=9975.7666  cpu=27.2193  opt=NPSOL
ver=2.19.5.1  stc=0

> fitEstCIs(fitACE1,colUS)   # this will give an error if you set intervals=F, but still print
the parameter estimates
interC  betaS  betaA   VA11   VC11   VE11
0.0746 0.0982 0.0016 0.4128 0.1907 0.2763
             VA     VC     VE     SA     SC     SE
lbound   0.3210 0.1019 0.2537 0.3656 0.1169 0.2858
estimate 0.4128 0.1907 0.2763 0.4692 0.2168 0.3140
ubound   0.5100 0.2764 0.3016 0.5774 0.3096 0.3452
```

7. What is the size of the predicted covariance matrix?
8. What proportion of the variance is due to additive genetic factors?

9. Run **Model 2** (lines 112-169) or open & run individual script 2: oneACEvc_2sib.R
10. Inspect the lines related to specifying the expected covariance matrices

```
# Create Algebra for expected Covariance Matrices in MZ & DZ twins & extra sibling
expCovMZ  <- mxAlgebra( expression= rbind( cbind(V,     cMZ,  cDZ),
                                           cbind(t(cMZ), V,   cDZ),
                                           cbind(t(cDZ), cDZ, V)), name="expCovMZ" )
expCovDZ  <- mxAlgebra( expression= rbind( cbind(V,      cDZ, cDZ),
                                           cbind(t(cDZ), V,   cDZ),
                                           cbind(t(cDZ), cDZ, V)), name="expCovDZ" )
```

11. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals

12. What is the size of the predicted covariance matrix?
13. What is the expected variance of siblings?
14. What about their covariance with their MZ or DZ co-twin?
15. Are parameter estimates and confidence intervals different from Model 1? Why?

16. Run **Model 3** (lines 170-206) or open & run individual script 3: oneACEvc_3alt.R
17. Inspect the lines related to specifying the expected covariance matrices

```
# Create Algebra for expected Covariance Matrices in MZ & DZ twins & extra sibling
relAmz    <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=FALSE, values=c(1,
 1,.5,1,.5,1), name="rAmz" )
relAdz    <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=FALSE,
 values=c(1,.5,.5,1,.5,1), name="rAdz" )
relC      <- mxMatrix( type="Unit", nrow=ntv, ncol=ntv, free=FALSE, name="rC" )
relE      <- mxMatrix( type="Iden", nrow=ntv, ncol=ntv, free=FALSE, name="rE" )
expCovMZ  <- mxAlgebra( expression= VA%x%rAmz + VC%x%rC + VE%x%rE, name="expCovMZ" )
expCovDZ  <- mxAlgebra( expression= VA%x%rAdz + VC%x%rC + VE%x%rE, name="expCovDZ" )
```

18. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
19. Write out the relationship coefficient matrices for MZs and DZs in matrix format.
20. What is the difference between the MZ & DZ groups?
21. Are parameter estimates and confidence intervals different from Model 2? Should they be?

22. Run **Model 4** (lines 207-237) or open & run individual script 4: oneACEvc_4def.R
23. Inspect the lines related to specifying the expected covariance matrices

```
# Create Algebra for expected Variance/Covariance Matrices in twins & extra sibling
relA      <- mxMatrix( type="Stand", nrow=ntv, ncol=ntv, free=FALSE,
 labels=c("data.relT","data.relS","data.relS"), name="rA" )
expCovTW  <- mxAlgebra( expression= VA%x%rA + VC%x%rC + VE%x%rE, name="expCovTW" )
```

24. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
25. Write out the relationship coefficient matrices for MZs and DZs in matrix format.
26. Are parameter estimates and confidence intervals different from Model 3? Should they be?

27. Run **Model 5** (lines 238-261) or open & run individual script 5: oneACEvc_5alt.R
28. Inspect the lines related to specifying the expected covariance matrices

```
# Create Algebra for expected Variance/Covariance Matrices in twins & extra sibling
relA      <- mxMatrix( type="Stand", nrow=ntv, ncol=ntv, free=FALSE,
 labels=c("data.rel12","data.rel13","data.rel23"), name="rA" )
```

29. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
30. Write out the relationship coefficient matrices for MZs and DZs in matrix format.
31. Are parameter estimates and confidence intervals different from Model 4?
32. Did it take longer to run? Why?

33. Run **Model 6** (lines 262-287) or open & run individual script 6: oneACEvc_6dzs.R
34. Inspect the lines related to specifying the data objects

```
# Create Data Objects for DZ Group
dataDZ     <- mxData( observed=tsDataS[tsDataS$zyg==2,], type="raw" )
```

35. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
36. Are parameter estimates and confidence intervals different from Model 5?
37. What happened to the confidence intervals?
38. Can model fit of this model be compared with previous models?

For Intermediate Users

1. Open oneACEvc7_2fill.R
2. Run **Model 1** (all lines up 111) after fixing missing piece in the following lines of code:

```
# Create Matrices for Variance Components
covA       <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svVa,
 label="VA11", name="VA" )
covC       <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svVa,
 label="VC11", name="VC" )
covE       <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svVe,
 label="VE11", name="VE" )

# Create Algebra for expected Variance/Covariance Matrices in MZ & DZ twins
covP       <- mxAlgebra( expression= VA+VC+VE, name="V" )
covMZ      <- mxAlgebra( expression= VA+VC, name="cMZ" )
covDZ      <- mxAlgebra( expression= VA+ VC, name="cDZ" )
expCovMZ   <- mxAlgebra( expression= rbind( cbind(V,       cMZ),
                                            cbind(t(cMZ), V)), name="expCovMZ" )
expCovDZ   <- mxAlgebra( expression= rbind( cbind(V,       cDZ),
                                            cbind(t(cDZ), V)), name="expCovDZ" )
```

3. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
4. What is the size of the predicted covariance matrix?
5. What proportion of the variance is due to additive genetic factors?

6. Run **Model 2** (lines 112-169) after fixing missing piece in the following lines of code:

```
# Create Algebra for expected Covariance Matrices in MZ & DZ twins & extra sibling
expCovMZ   <- mxAlgebra( expression= rbind( cbind(V,      cMZ,   cDZ),
                                            cbind(t(cMZ), V,    cDZ),
                                            cbind(t(cDZ), cDZ, V)), name="expCovMZ" )
expCovDZ   <- mxAlgebra( expression= rbind( cbind(V,       cDZ, cDZ),
                                            cbind(t(cDZ), V,    cDZ),
                                            cbind(t(cDZ), cDZ, V)), name="expCovDZ" )
```

7. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
8. What is the size of the predicted covariance matrix?
9. What is the expected variance of siblings?
10. What about their covariance with their MZ or DZ co-twin?
11. Are parameter estimates and confidence intervals different from Model 1? Why?

12. Run **Model 3** (lines 170-206) after fixing missing piece in the following lines of code:

```
# Create Algebra for expected Covariance Matrices in MZ & DZ twins & extra sibling
relAmz    <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=FALSE, values=c(1,
 1,.5,1,.5,1), name="rAmz" )
relAdz    <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=FALSE,
 values=c(1,,,1,.5,1), name="rAdz" )
relC      <- mxMatrix( type="Unit", nrow=ntv, ncol=ntv, free=FALSE, name="rC" )
relE      <- mxMatrix( type="Iden", nrow=ntv, ncol=ntv, free=FALSE, name="rE" )
expCovMZ  <- mxAlgebra( expression= VA%x%rAmz + VC%x%rC + VE%x%rE, name="expCovMZ" )
expCovDZ  <- mxAlgebra( expression= VA%x%rAdz + VC%x%rC + VE%x%rE, name="expCovDZ" )
```

13. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
14. Write out the relationship coefficient matrices for MZs and DZs in matrix format.
15. What is the difference between the MZ & DZ groups?
16. Are parameter estimates and confidence intervals different from Model 2? Should they be?

17. Run **Model 4** (lines 207-237) after fixing missing piece in the following lines of code:

```
# Create Algebra for expected Variance/Covariance Matrices in twins & extra sibling
relA      <- mxMatrix( type="Stand", nrow=ntv, ncol=ntv, free=FALSE,
 labels=c("data.","data.relS","data.relS"), name="rA" )
expCovTW  <- mxAlgebra( expression= VA%x%rA + VC%x%rC + VE%x%rE, name="expCovTW" )
```

18. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
19. Write out the relationship coefficient matrices for MZs and DZs in matrix format.
20. Are parameter estimates and confidence intervals different from Model 3? Should they be?

21. Run **Model 5** (lines 238-261) after fixing missing piece in the following lines of code:

```
# Create Algebra for expected Variance/Covariance Matrices in twins & extra sibling
relA      <- mxMatrix( type="", nrow=ntv, ncol=ntv, free=FALSE,
 labels=c("data.rel12","data.rel13","data.rel23"), name="rA" )
```

22. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
23. Are parameter estimates and confidence intervals different from Model 4?
24. Did it take longer to run? Why?
25. Does using the actual relatedness impact the running time of the script?

26. Run **Model 6** (lines 262-287) after fixing missing piece in the following lines of code:

```
# Create Data Objects for DZ Group
dataDZ    <- mxData( observed=tsDataS[tsDataS$zyg==,], type="raw" )
```

27. Record goodness-of-fit (df, -2 log-likelihood), parameter estimates and confidence intervals
28. Are parameter estimates and confidence intervals different from Model 5?
29. What happened to the confidence intervals?
30. Can model fit of this model be compared with previous models?

Answer Sheet

Goodness-of-fit statistics									
	os	ns	ep	co	df	ll	cpu	stc	
1cov	4000	2000	6	0	3994	9975.76	19.761	0	
2sib	6000	2000	6	0	5994	14873.6	21.7278	0	
3alt	6000	2000	6	0	5994	14873.6	21.6779	0	
4def	6000	2000	6	0	5994	14873.6	37.9076	0	
5rel	6000	2000	6	0	5994	14874.5	42.0952	0	
6dzs	3000	1000	6	0	2994	7589.48	16.4638	0	
Parameter estimates									
	interC	betaS	betaA	VA11	VC11	VE11			
1cov	0.0746	0.0982	0.0016	0.4128	0.1907	0.2763			
2sib	0.1475	0.0566	0.0002	0.4383	0.1711	0.277			
3alt	0.1475	0.0566	0.0002	0.4383	0.1711	0.277			
4def	0.1475	0.0566	0.0002	0.4383	0.1711	0.277			
5rel	0.1471	0.0571	0.0002	0.4368	0.1725	0.2778			
6dzs	0.1551	0.0641	-0.0003	0.1252	0.3201	0.4294			
Standardized variance components and their confidence intervals									
	lbA	**pA**	ubA	lbC	**pC**	ubC	lbE	**pE**	ubE
1cov	0.3656	**0.4692**	0.5774	0.1169	**0.2168**	0.3096	0.2858	**0.314**	0.3452
2sib	0.4238	**0.4945**	0.5643	0.1349	**0.1931**	0.2496	0.2849	**0.3125**	0.3429
3alt	0.4238	**0.4945**	0.5643	0.1349	**0.1931**	0.2496	0.2849	**0.3125**	0.3429
4def	0.4238	**0.4945**	0.5643	0.1349	**0.1931**	0.2496	0.2849	**0.3125**	0.3429
5rel	0.4216	**0.4924**	0.5623	0.1363	**0.1945**	0.251	0.2855	**0.3132**	0.3437
6dzs	-0.6722	**0.1431**	0.9498	-0.04	**0.366**	0.77	0.0894	**0.4909**	0.903

os: number of observed statistics
ns: number of records
ep: estimated parameters
df: degrees of freedom
ll: -2 log-likelihood of data
cpu: computer processing time
etc: status code (0 is ok)

lb: lower bound of confidence interval
ub: upper bound of confidence interval
pA: proportion of variance explained by additive genetic factors
pC: proportion of variance explained by shared environmental factors
pE: proportion of variance explained by unique environmental factors

For Advanced Users

1. Fix & Run oneACEvc7_2fill.R
2. Simulate new data using simTSc.R with different sample size and/or expected proportions of variance for ACE
3. Evaluate sample size needed to get similar confidence intervals with DZ only model to model with both MZ & DZ twins, possibly under various scenarios of expected proportions

Answer Key for Beginners

7. What is the size of the predicted covariance matrix? 2x2
8. What proportion of the variance is due to additive genetic factors? 47%

12. What is the size of the predicted covariance matrix? 3x3
13. What is the expected variance of siblings? A+C+E
14. What about their covariance with their MZ or DZ co-twin? 0.5A+C
15. Are parameter estimates and confidence intervals different from Model 1? Why? maybe, if sibling correlation is not exactly the same as the DZ correlation

19. Write out the relationship coefficient matrices for MZs and DZs in matrix format. See video
20. What is the difference between the MZ & DZ groups? 1 & .5
21. Are parameter estimates and CIs different from Model 2? Should they be? No

25. Write out the relationship coefficient matrices for MZs and DZs in matrix format. See video
26. Are parameter estimates and CIs different from Model 3? Should they be? No

30. Write out the relationship coefficient matrices for MZs and DZs in matrix format. See video
31. Are parameter estimates and CIs different from Model 4? No
32. Did it take longer to run? Why? Yes, to read in definition variables

36. Are parameter estimates and confidence intervals different from Model 5? Yes
37. What happened to the confidence intervals? They exploded
38. Can model fit of this model be compared with previous models? Not really - different data

Answer Key for Intermediate (different numbering)

4. What is the size of the predicted covariance matrix? 2x2
5. What proportion of the variance is due to additive genetic factors? 47%

8. What is the size of the predicted covariance matrix? 3x3
9. What is the expected variance of siblings? A+C+E
10. What about their covariance with their MZ or DZ co-twin? 0.5A+C
11. Are parameter estimates and confidence intervals different from Model 1? Why? maybe, if sibling correlation is not exactly the same as the DZ correlation

14. Write out the relationship coefficient matrices for MZs and DZs in matrix format. See video
15. What is the difference between the MZ & DZ groups? 1 & .5
16. Are parameter estimates and CIs different from Model 2? Should they be? No

19. Write out the relationship coefficient matrices for MZs and DZs in matrix format. See video
20. Are parameter estimates and CIs different from Model 3? Should they be? No

23. Are parameter estimates and CIs different from Model 4? No
24. Did it take longer to run? Why? Yes, to read in definition variables
25. Does using the actual relatedness impact the running time of the script? Slows down

28. Are parameter estimates and confidence intervals different from Model 5? Yes
29. What happened to the confidence intervals? They exploded
30. Can model fit of this model be compared with previous models? Not really - different data