Introduction to Analysis of Twin Data Using R and OpenMx- Part 2

Elizabeth Prom-Wormley

Special Thanks to

Sarah Medland, Hermine Maes,

Nick Martin, and Lucia Colodro Conde

Files for this session, go to:
/home/elizabeth/2022/

Copy 4 files to your server folder:
IntroToTwinAnalysisInRAndOpenMxMay2022.R
miFunctions.R
SimWtDataInd.csv
SimWtDataPair.csv

Session Objectives

By the end of the session, you will be able to

- Summarize the goals of analyzing twin data for a single phenotype
- Summarize the general process by which to analyze twin data for a single phenotype
- Apply basic R functions for twin-focused data analysis and visualization
- Translate the implementation of a basic ACE model between a structural equation model and an OpenMx model

Analysis of Twin Data
Overall Questions to be Answered

- Does a trait of interest cluster among related individuals ?
- Can clustering be explained by genetic and/or environmental effects?
- Which sources of variance contribute significantly to the total variance of the trait?
- How much of the variation in the trait is due to genetic and environmental effects?

Classical Twin Studies
Basic Background

- The Classical Twin Study (CTS) uses MZ and DZ twins reared together
  - MZ twins share 100% of their genes
  - DZ twins share *on average* 50% of their genes

- Expectation- Genetic factors are assumed to contribute to a phenotype when MZ twins are more similar than DZ twins

A Roadmap for Analysis of A Single Phenotype (Univariate or Monophenotype) Analysis

1. Estimate basic summary statistics and visualize data
   - Means, variances, covariances, correlations by zygosity group
2. Test basic data assumptions (equal means & variances for twin 1/twin 2 and MZ/DZ pairs)
   - *Saturated Model*
3. Estimate contributions of genetic and environmental effects on the total variance of a phenotype
   - *ACE or ADE Models*
4. Test ACE (or ADE) submodels to identify and report significant genetic and environmental contributions
   - *AE or CE or E Only Models*

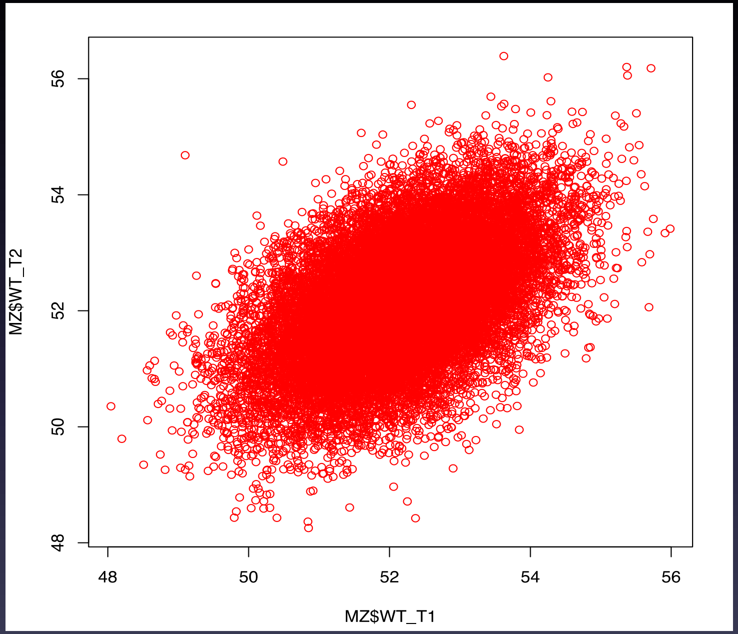
A Quick Look at the Data

Study Sample

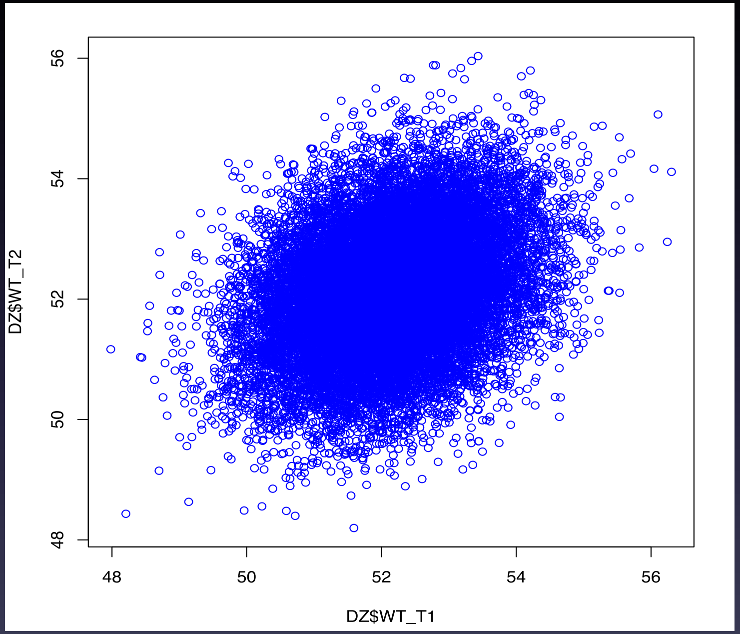
Simulated data of weight (in kilograms) in female adults 65 and older

N = 40,000 pairs
MZ =20,000
DZ = 20,000

The study was conducted in two locations (study sites)



rMZ = 0.52

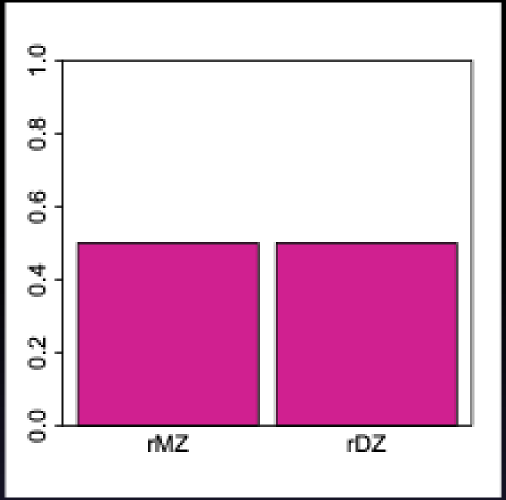


rMZ = 0.31

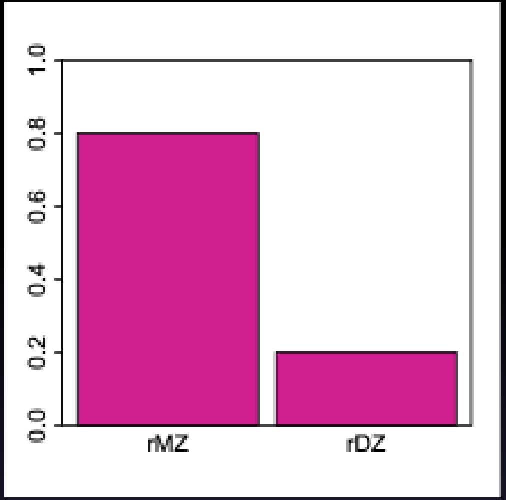
Patterns of Twin Correlation

rMZ = 2rDZ
Additive

DZ twins on average share 50% of additive effects

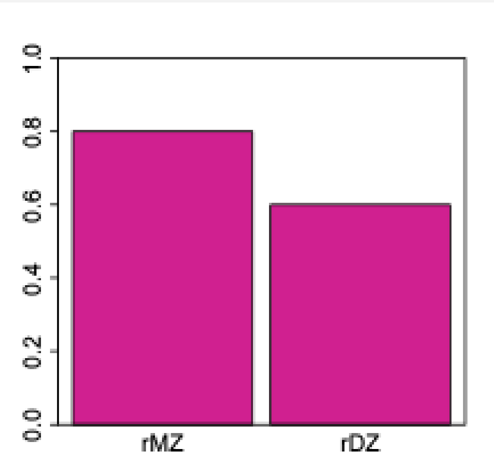


rMZ = rDZ
Shared Environment



rMZ > 2rDZ
Additive & Dominance

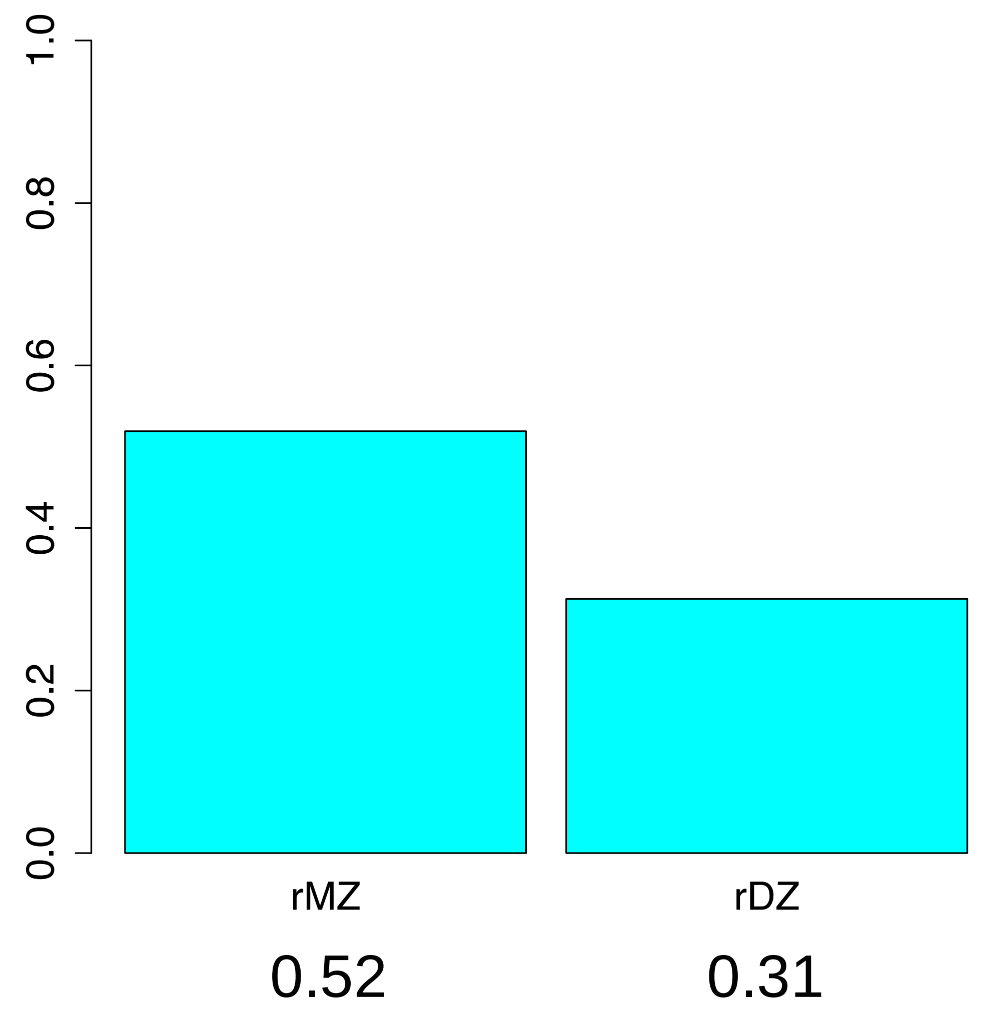
DZ twins on average share 25% of dominance effects



Additive & “Shared Environment”

A = 2(rMZ-rDZ)
C = 2rDZ - rMZ
E = 1- rMZ

Twin Correlations



A = 2(rMZ-rDZ)

C = 2rDZ - rMZ

E = 1- rMZ

A = 0.42

C = 0.10

E = 0.58

ADE or ACE?

“Old Fashioned” Data Checking

	MZ		DZ	
	T1	T2	T1	T2
mean	52.157	52.154	52.154	52.152
variance	1.05	1.03	1.019	1.016
covariance$_{(T1-T2)}$	0.54		0.32	

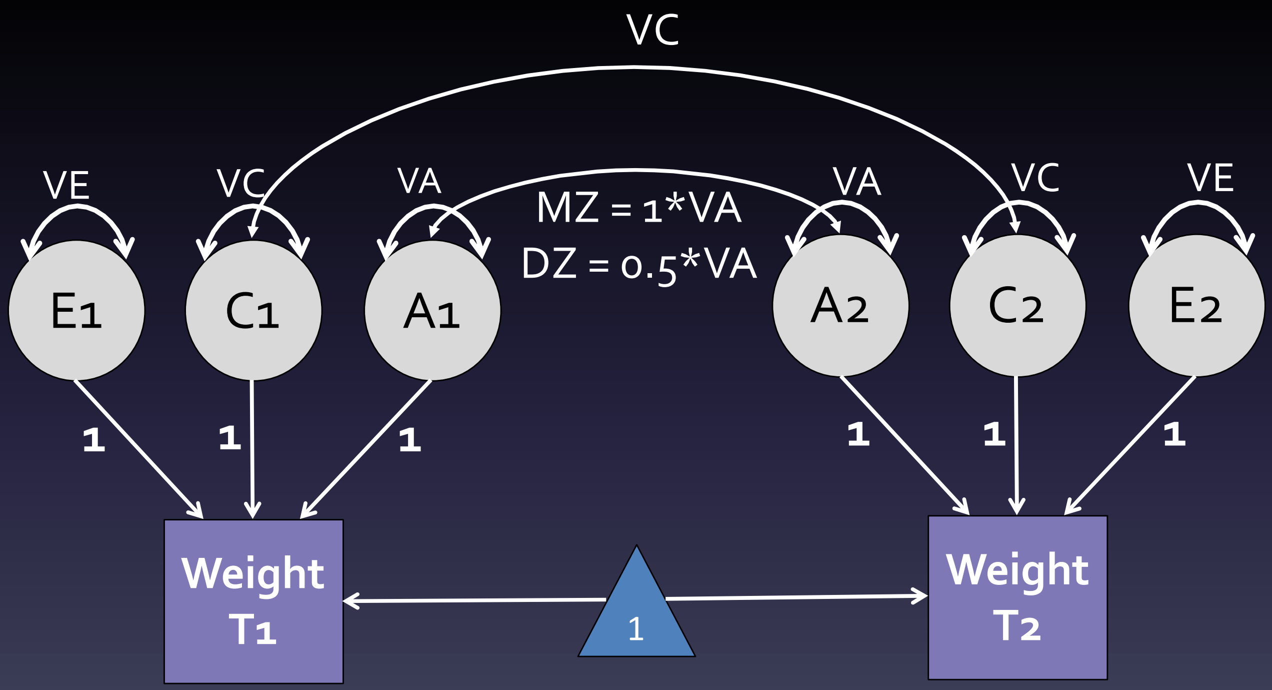
This looks very convincing, but how can we be sure that these means and variances are truly the same?

A Roadmap for Analysis of A Single Phenotype (Univariate or Monophenotype) Analysis

1. Estimate basic summary statistics and visualize data
   - Means, variances, covariances, correlations by zygosity group
2. Test basic data assumptions (equal means & variances for twin 1/twin 2 and MZ/DZ pairs)
   - *Saturated Model*
3. Estimate contributions of genetic and environmental effects on the total variance of a phenotype
   - *ACE or ADE Models*
4. Test ACE (or ADE) submodels to identify and report significant genetic and environmental contributions
   - *AE or CE or E Only Models*

Classical Twin Studies
Data Assumptions

- MZ and DZ twins are sampled from the same population, therefore we expect:
  - Equal means/variances in Twin 1 and Twin 2
  - Equal means/variances in MZ and DZ twins
- Further assumptions would need to be tested if we introduce male twins and opposite sex twin pairs



Fitting Nested Models

- Saturated Model
  - Likelihood of data without any constraints
  - Fitting as many means, variances, and covariances as possible

- Equality of means & variances by twin order
  - Test if mean of twin 1 = mean of twin 2
  - Test if variance of twin 1 = variance of twin 2

- Equality of means & variances by zygosity
  - Test if mean of MZ = mean of DZ
  - Test if variance of MZ = variance of DZ

Estimated Values From Saturated Model

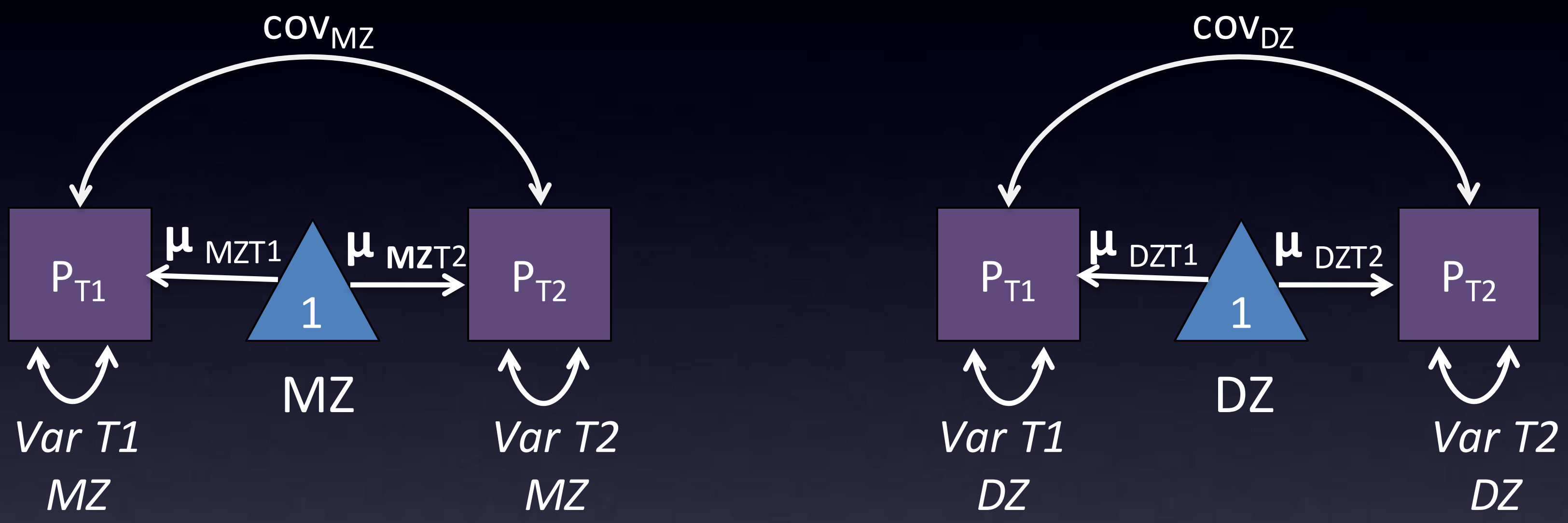
		T1	T2		T1	T2
Saturated Model						
mean	MZ	52.1573	52.1543	DZ	52.1536	52.1518
cov	T1	1.0454	X	T1	1.0189	X
	T2	0.5399	1.0348	T2	0.3182	1.0158

10 Total Parameters Estimated

mMZ1, mMZ2,
vMZ1,vMZ2,cMZ21
mDZ1, mDZ2,
vDZ1,vDZ2,cDZ21

Standardize covariance matrices (covMZ & covDZ) to calculate twin pair correlations (rMZ & rDZ)

Saturated Twin Model



10 Total Parameters Estimated

mMZ1, mMZ2,
vMZ1,vMZ2,cMZ21
mDZ1, mDZ2,
vDZ1,vDZ2,cDZ21

Saturated Model Fit Comparisons

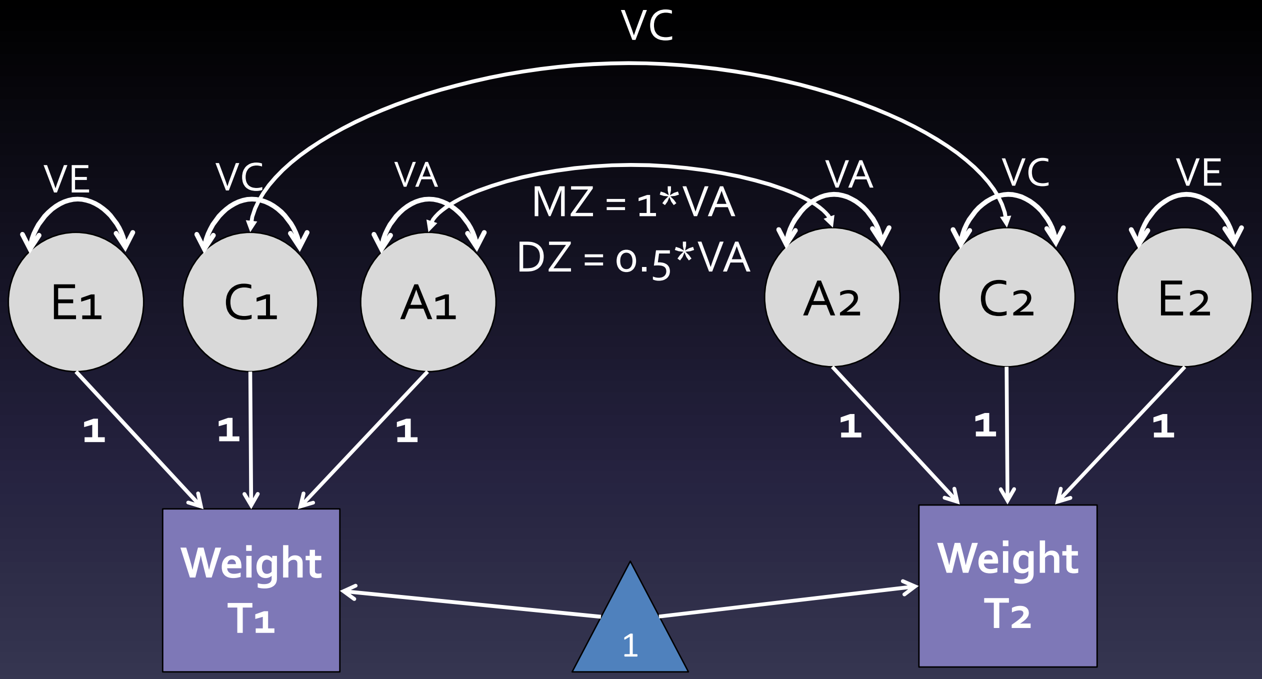
Model	ep	-2ll	df	AIC	diff -2ll	diff df	p
Saturated	10	220950.05	79990	60970.048			
mT1=mT2	8	220950.27	79992	60966.273	0.22	2	0.89
mT1=mT2 & varT1=VarT2	6	220951.04	79994	60963.037	0.99	4	0.91
Zyg MZ=DZ	4	220955.31	79996	60963.312	5.26	6	0.51

No significant differences between saturated model and models where means/variances/covariances are equal by zygosity and between twins.

A Roadmap for Analysis of A Single Phenotype (Univariate or Monophenotype) Analysis

1. Estimate basic summary statistics and visualize data
   - Means, variances, covariances, correlations by zygosity group
2. Test basic data assumptions (equal means & variances for twin 1/twin 2 and MZ/DZ pairs)
   - *Saturated Model*
3. Estimate contributions of genetic and environmental effects on the total variance of a phenotype
   - *ACE or ADE Models*
4. Test ACE (or ADE) submodels to identify and report significant genetic and environmental contributions
   - *AE or CE or E Only Models*

ACE Model



This model has 4 estimated parameters:

1- Mean of weight
2- Variance due to A (VA)
3- Variance due to C (VC)
4- Variance due to E (VE)

ACE Model Assumptions

- MZ twins are genetically identical
- Equal Environments of MZ and DZ pairs
- Random Mating
- Genetic and environmental influences are independent and additive
    - No Gene-Environment Correlation (rGE)
    - No Gene-Environment Interaction (GxE)
    - No Sex Limitation
    - No Gene-Age Interaction
    - No Epistasis

ACE Model Fit Comparisons

Model	ep	-2ll	df	AIC	diff -2ll	diff df	p
ACE	4	220955.31	79996	60963.31			
AE	3	221028.80	79997	61034.80	73.49	1	<1e-15
CE	3	221650.71	79997	61656.71	695.39	1	<1e-15
E Only	2	229295.31	79998	69299.31	8339.99	2	<1e-15

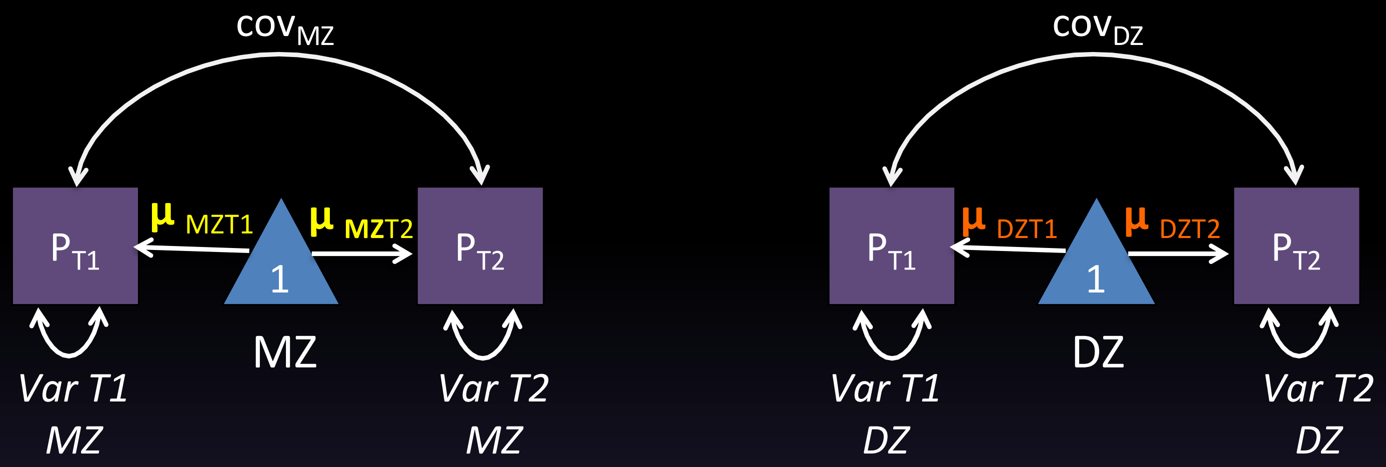
An ACE model best describes the data

Parameter Estimates from the best fitting model

VA	VC	VE	SA	SC	SE
0.4115	0.1185	0.4979	0.4004	0.1153	0.4844

Want details on implementing the code used to produce the results used here? Keep watching!

Saturated Code Deconstructed



```
meanMZ <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE,
values=svMe, labels=c("mMZ1","mMZ2"), name="meanMZ" )
```

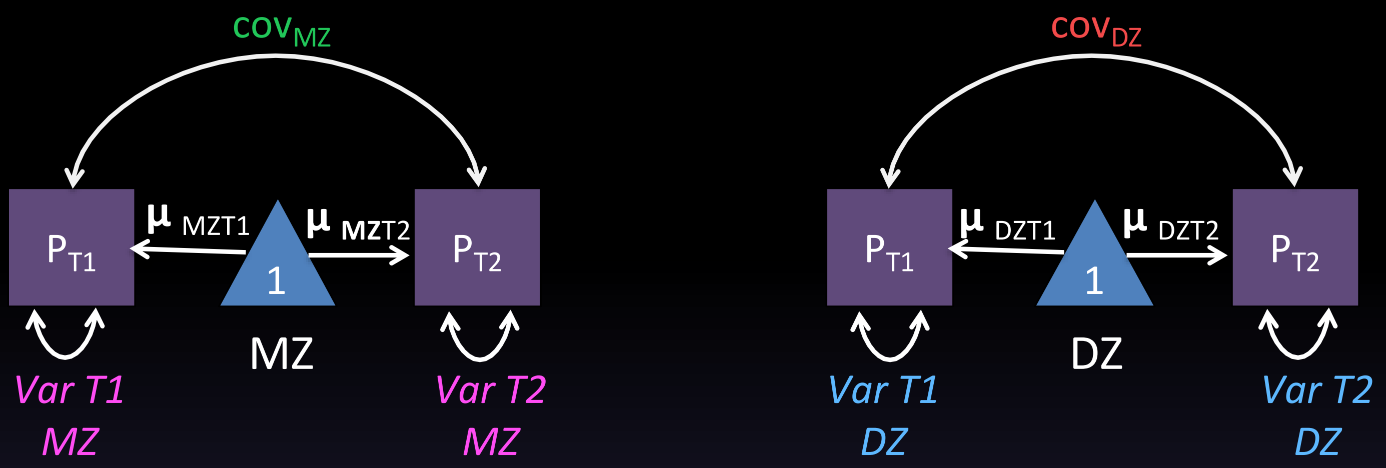
mMZ1	mMZ2

```
meanDZ <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE,
values=svMe, labels=c("mDZ1","mDZ2"), name="meanDZ" )
```

mDZ1	mDZ2

mean DZ = 1 x 2 matrix

Saturated Code Deconstructed



```
covMZ    <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv,
free=TRUE, values=valDiag(svVa,ntv), lbound=valDiag(lbVa,ntv),
labels=c("vMZ1","cMZ21","vMZ2"), name="expCovMZ" )
```

	T1	T2
T1	vMZ1	cMZ21
T2	cMZ21	vMZ2

covMZ = 2 x 2 matrix

```
covDZ    <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv,
free=TRUE, values=cvVals, lbound=lbVals,
labels=c("vDZ1","cDZ21","vDZ2"), name="expCovDZ" )
```

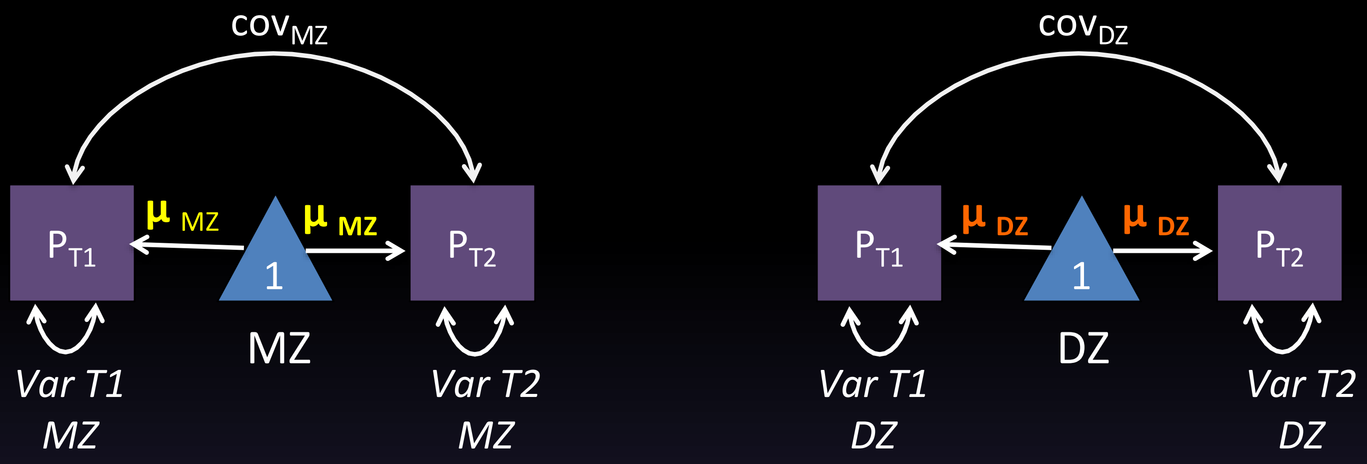
	T1	T2
T1	vDZ1	cDZ21
T2	cDZ21	vDZ2

covDZ = 2 x 2 matrix

```
# Create Data Objects for Multiple Groups
dataMZ <- mxData( observed=mzData, type="raw" )
dataDZ <- mxData( observed=dzData, type="raw" )

# Create Expectation Objects for Multiple Groups
expMZ <- mxExpectationNormal( covariance="covMZ", means="meanMZ", dimnames=selVars )
expDZ <- mxExpectationNormal( covariance="covDZ", means="meanDZ", dimnames=selVars )
funML <- mxFitFunctionML()

# Create Model Objects for Multiple Groups
modelMZ <- mxModel( meanMZ, covMZ, dataMZ, expMZ, funML, name="MZ" )
modelDZ <- mxModel( meanDZ, covDZ, dataDZ, expDZ, funML, name="DZ" )
multi <- mxFitFunctionMultigroup( c("MZ","DZ") )

# Create Confidence Interval Objects
ciCov <- mxCI( c('MZ.covMZ','DZ.covDZ') )
ciMean <- mxCI( c('MZ.meanMZ','DZ.meanDZ') )
```

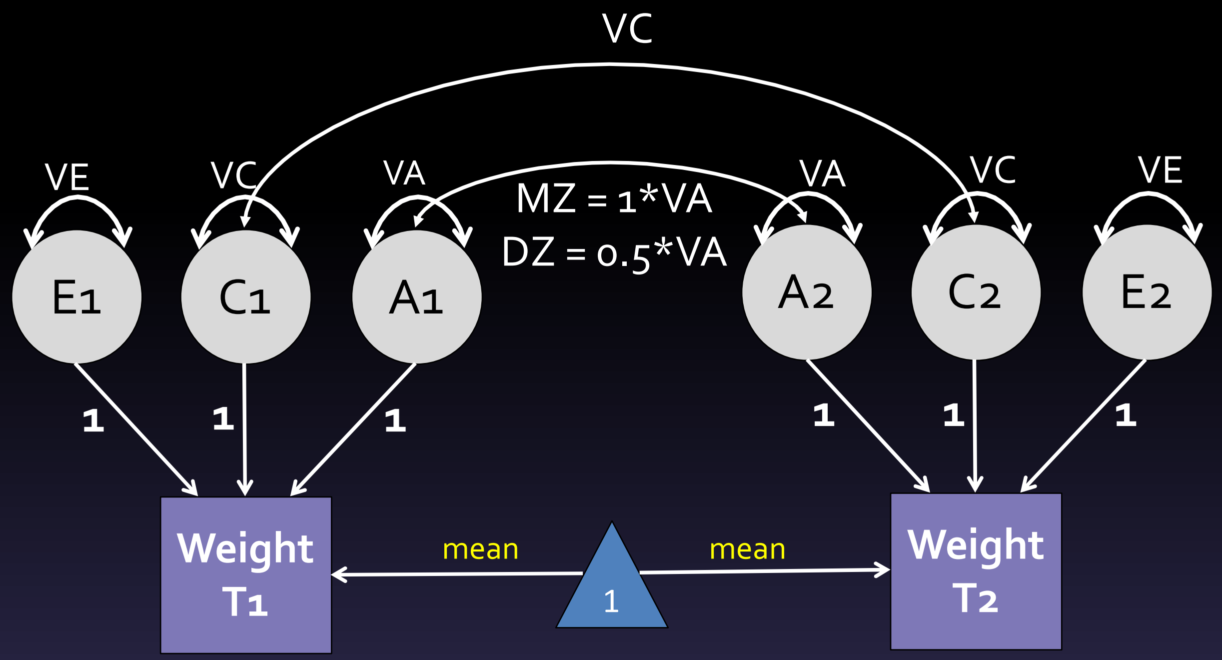
```
# Build Saturated Model with Confidence Intervals
modelSAT <- mxModel( "oneSATc", modelMZ, modelDZ, multi, ciCov, ciMean )

# Run Saturated Model
fitSAT <- mxRun( modelSAT, intervals=F )
sumSAT <- summary( fitSAT )
```

Submodel 1- Equal Means Across Twins



```
modelEMO <- mxModel( fitSAT, name="oneEMOc" )

modelEMO <- omxSetParameters( modelEMO, label=c("mMZ1","mMZ2"),
free=TRUE, values=svMe, newlabels='mMZ’ )

modelEMO <- omxSetParameters( modelEMO, label=c("mDZ1","mDZ2"),
free=TRUE, values=svMe, newlabels='mDZ' )
```

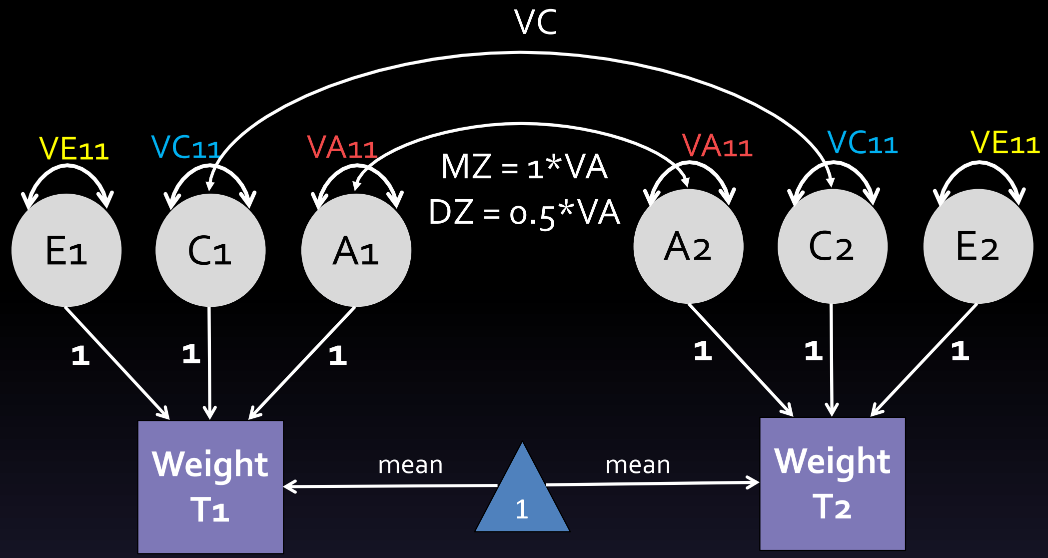
ACE Model Deconstructed- Means



```
meanG <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE,
          values=svMe, labels=labVars("mean",vars), name="meanG" )
```

T1	T2
mean	mean

ACE Model Deconstructed- Variances



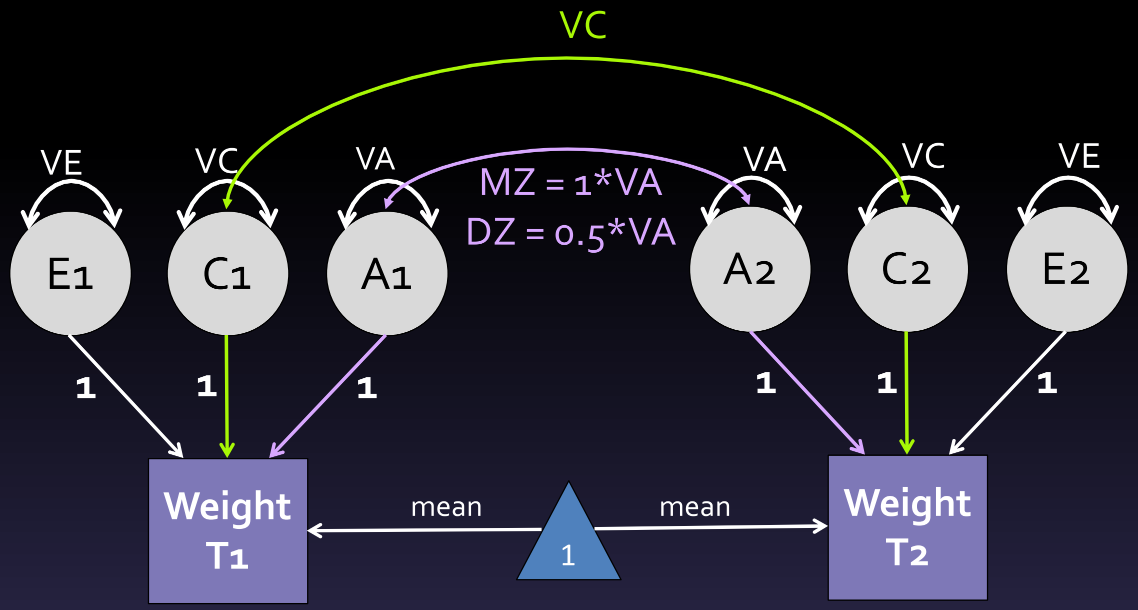
```
covA <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svPa,
          label="VA11", name="VA" )

covC <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svPa,
          label="VC11", name="VC" )

covE <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svPa,
          label="VE11", name="VE" )
```

VA11 *1 x 1 matrix*

VC11 *1 x 1 matrix*

VE11 *1 x 1 matrix*

ACE Model Deconstructed Variances/Covariances



```
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" )
```

MZ Twins

	T1	T2
T1	VA + VC +VE	VA + VC
T2	VA + VC	VA + VC+VE

DZ Twins

	T1	T2
T1	VA + VC + VE	0.5VA + VC
T2	0.5VA + VC	VA + VC + VE

```
# Create Data Objects for Multiple Groups
dataMZ <- mxData( observed=mzData, type="raw" )
dataDZ <- mxData( observed=dzData, type="raw" )

# Create Expectation Objects for Multiple Groups
expMZ <- mxExpectationNormal( covariance="expCovMZ", means="meanG", dimnames=selVars )
expDZ <- mxExpectationNormal( covariance="expCovDZ", means="meanG", dimnames=selVars )
funML <- mxFitFunctionML()

# Create Model Objects for Multiple Groups
pars <- list( meanG, covA, covC, covE, covP )
modelMZ <- mxModel( pars, covMZ, expCovMZ, dataMZ, expMZ, funML, name="MZ" )
modelDZ <- mxModel( pars, covDZ, expCovDZ, dataDZ, expDZ, funML, name="DZ" )
multi <- mxFitFunctionMultigroup( c("MZ","DZ") )

# Create Algebra for Unstandardized and Standardized Variance Components
rowUS <- rep('US',nv)
colUS <- rep(c('VA','VC','VE','SA','SC','SE'),each=nv)
estUS <- mxAlgebra( expression=cbind(VA,VC,VE,VA/V,VC/V,VE/V), name="US", dimnames=list(rowUS,colUS) )
```

```
# Build Model with Confidence Intervals
modelACE <- mxModel( "oneACEvc", pars, modelMZ, modelDZ, multi, estUS, ciACE )



# Run ACE Model
fitACE <- mxRun( modelACE, intervals=TRUE )
sumACE <- summary( fitACE )
```

Thank You!

Questions?

Elizabeth.Prom-Wormley@vcuhealth.org