

Univariate Twin Analysis

OpenMx Tc26 2012

Hermine Maes, Elizabeth Prom-Wormley, Nick
Martin

Answers to be

Answered

Does a trait of interest cluster among related individuals?

Can clustering be explained by genetic or environmental effects?

What is the best way to explain the degree to which genetic and environmental effects affect a trait?

Family & Twin Study Designs

Family Studies

Classical Twin Studies

Adoption Studies

The Data

Australian Twin Register

18-30 years old, males and females

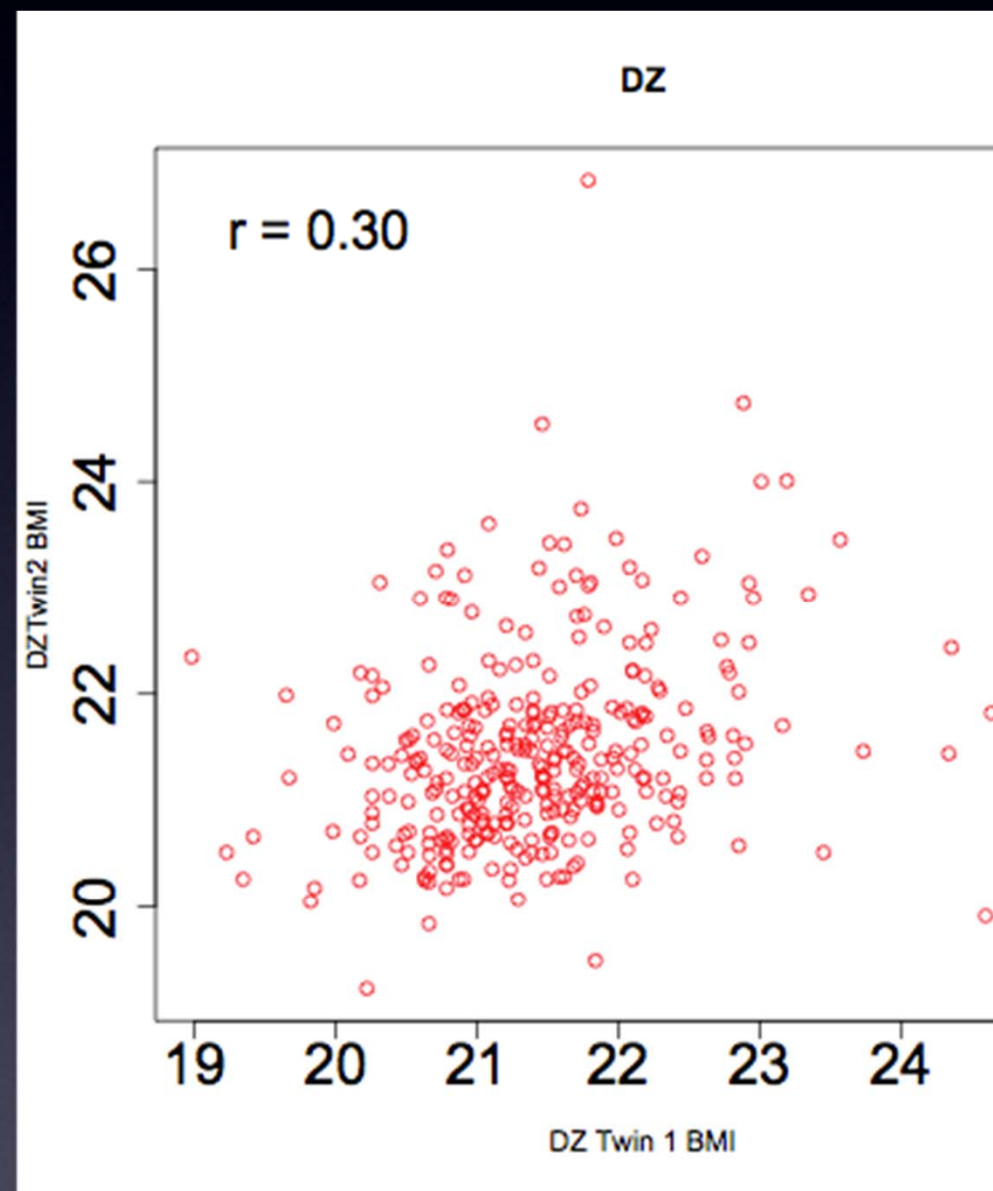
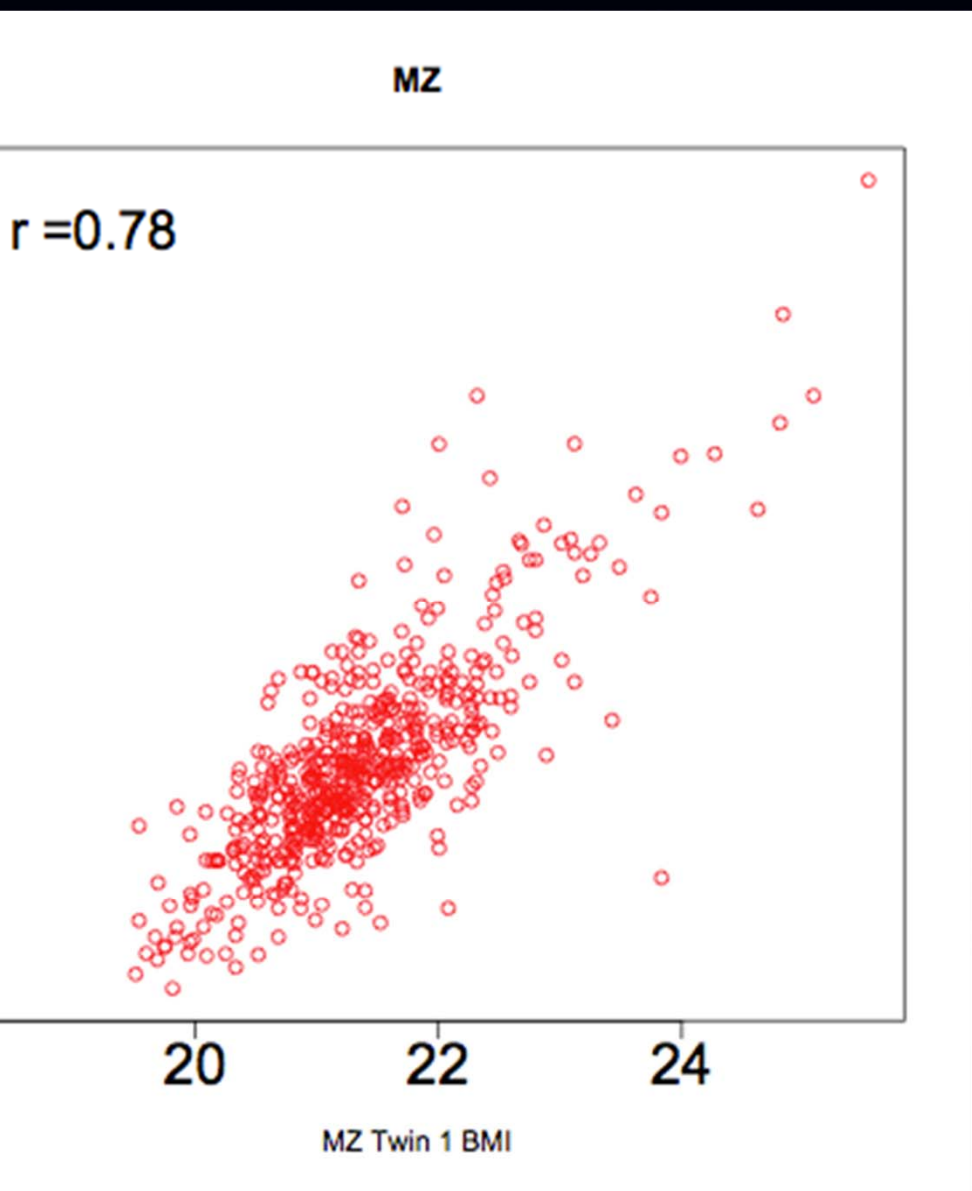
Work from this session will focus on Body Mass Index ($\text{weight}/\text{height}^2$) in females only

Sample size

MZF = 534 complete pairs (zyg = 1)

DZF = 328 complete pairs (zyg = 3)

A QUICK LOOK at the Data



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 twi

Classical Twin Study

Assumptions

- MZ twins are genetically identical
- Equal Environments of MZ and DZ pairs

Basic 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 twin pairs

Further assumptions would need to be tested if we introduce male twins and opposite sex twin pairs

Checking

	MZ		DZ	
	T1	T2	T1	T2
mean	21.35	21.34	21.45	21.46
variance	0.73	0.79	0.77	0.82
variance _(T1-T2)	0.59		0.25	

ce, but how can we actually be sure that the means and variances are truly the same?

A Roadmap

Use the data to test basic assumptions (equality of means & variances for twin 1/twin 2 and MZ/DZ pairs)

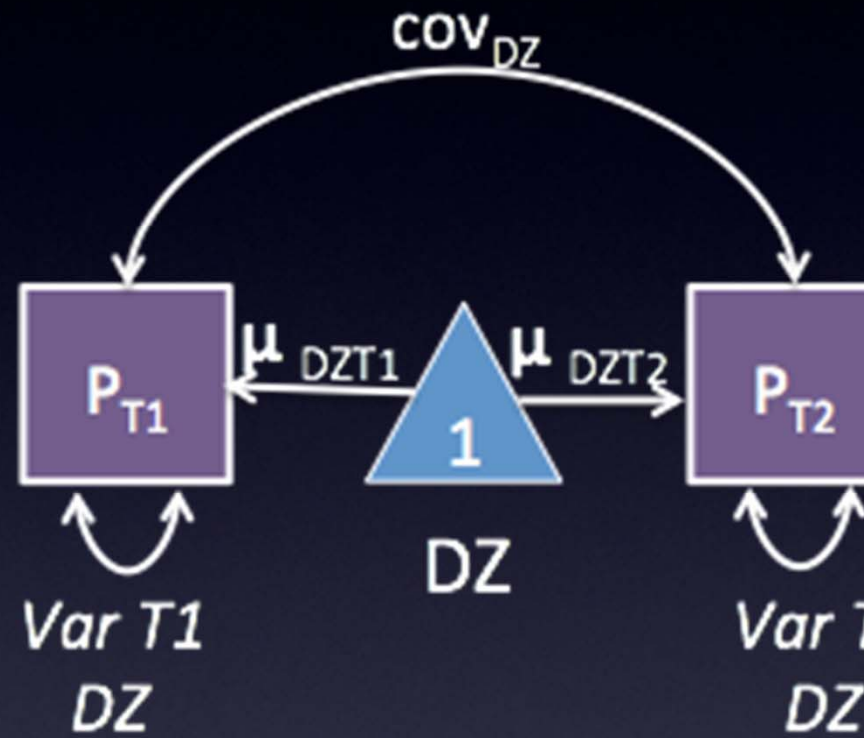
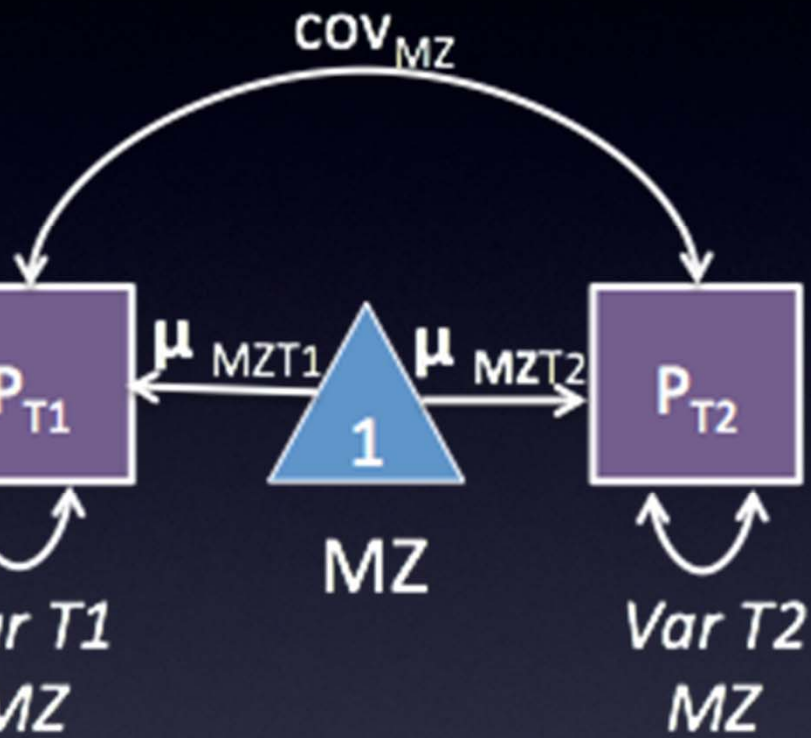
Saturated Model

Estimate contributions of genetic and environmental effects on the total variance of phenotype

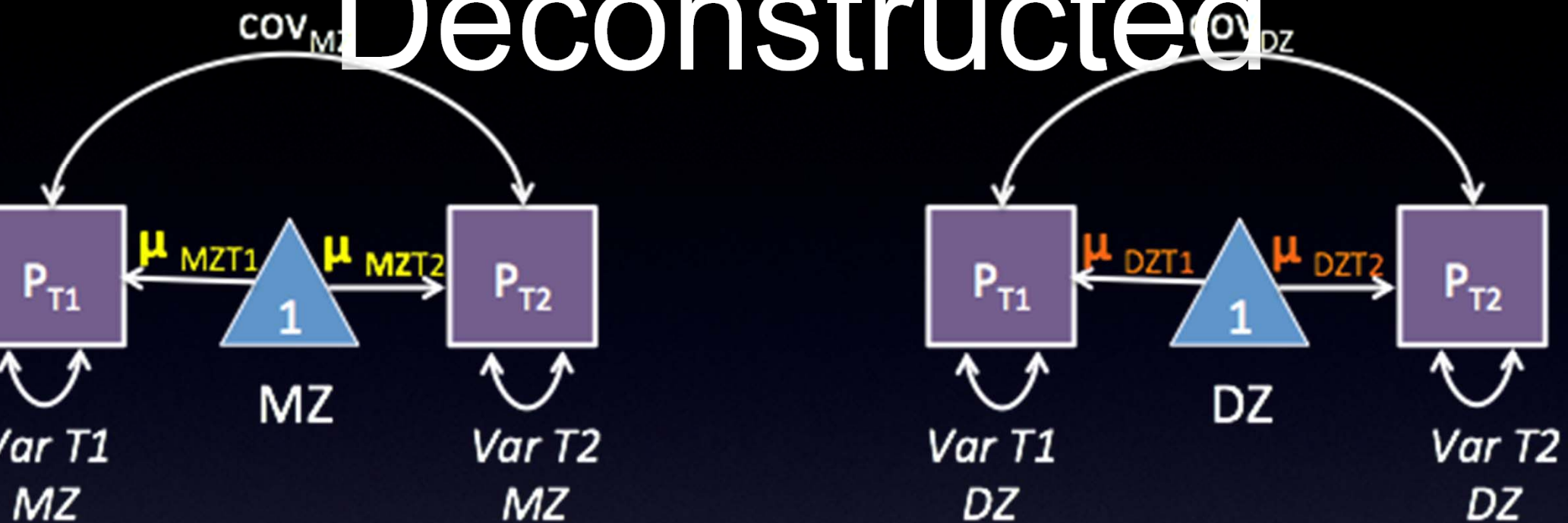
ACE or ADE Models

Test ACE (ADE) submodels to identify and report significant genetic and environmental

Saturated Twin Model



Deconstructed



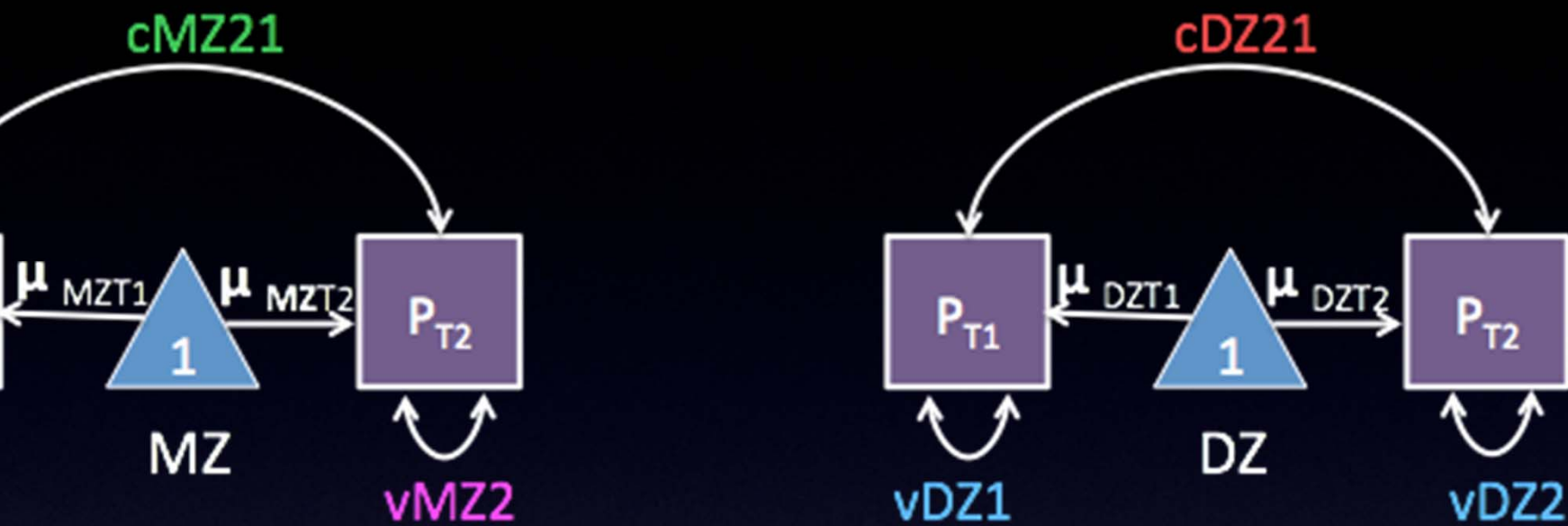
```
meanMZ <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE,  
  values=meVals, labels=c("mMZ1", "mMZ2"), name="expMeanMZ" )
```

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

mMZ1	mMZ2
------	------

mDZ1	mDZ2
------	------

Latent Class Growth Models



```

<- mxMatrix( type="Symm", nrow=ntv, ncol=ntv,
RUE, values=cvVals, lbound=lbVals, labels=c
", "cMZ21", "vMZ2"), name="expCovMZ" )

```

	T1
T1	vMZ1
T2	cMZ21

covMZ = 2

```

<- mxMatrix( type="Symm", nrow=ntv, ncol=ntv,
RUE, values=cvVals, lbound=lbVals, labels=c
", "cDZ21", "vDZ2"), name="expCovDZ" )

```

	T1
T1	vDZ1
T2	cDZ21

Time to Play...
with OpenMx

Please Open the File
twinSatCon.R

Estimated values

		T1	T2		T1	T2
Saturated Model						
mean	MZ			DZ		
cov	T1			T1		
	T2			T2		

0 Total Parameters

μ_{MZ1} , μ_{MZ2} ,
estimated

σ_{MZ1} , σ_{MZ2} , ρ_{MZ21}

μ_{DZ1} , μ_{DZ2} ,

σ_{DZ1} , σ_{DZ2} , ρ_{DZ21} Variance matrices for

in pair correlations (cov_{MZ} &

Estimated values

		T1	T2		T1	T2
Saturated Model						
mean	MZ	21.34	21.35	DZ	21.45	21.46
cov	T1	0.73		T1	0.77	
	T2	0.59	0.79	T2	0.24	0.82

0 Total Parameters

μ_{MZ1} , μ_{MZ2} ,
 estimated
 σ_{MZ1} , σ_{MZ2} , c_{MZ21}

μ_{DZ1} , μ_{DZ2} ,

σ_{DZ1} , σ_{DZ2} , c_{DZ21} Variance matrices for

in pair correlations (cov_{MZ} &

Fitting Nested Models

- Saturated Model
 - likelihood of data without any constraints
 - fitting as many means and (co)variances 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

		T1	T2		T1	T2
Equate Means & Variances across Twin Order						
mean	MZ			DZ		
cov	T1			T1		
	T2			T2		

Equate Means Variances across Twin Order & Zygosity

mean	MZ			DZ		
cov	T1			T1		
	T2			T2		

Estimates

		T1	T2		T1	T2
Equate Means & Variances across Twin Order						
mean	MZ	21.35	21.35	DZ	21.45	21.45
cov	T1	0.76		T1	0.79	
	T2	0.59	0.76	T2	0.24	0.79

Equate Means Variances across Twin Order & Zygosity

mean	MZ	21.39	21.39	DZ	21.39	21.39
cov	T1	0.78		T1	0.78	
	T2	0.61	0.78	T2	0.23	0.78

Stats

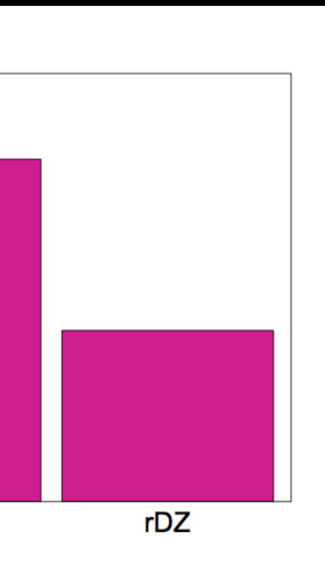
Model	ep	-2ll	df	AIC	diff - 2ll	diff df	
Saturated	10	4055.93	1767	521.93			
mT1=mT2	8	4056	1769	518	0.07	2	0
mT1=mT2 & varT1=VarT2	6	4058.94	1771	516.94	3.01	4	0
Zyg MZ=DZ	4	4063.45	1773	517.45	7.52	6	0

significant differences between saturated models
models where means/variances/covariance
equal by zygosity and between twins

Questions?

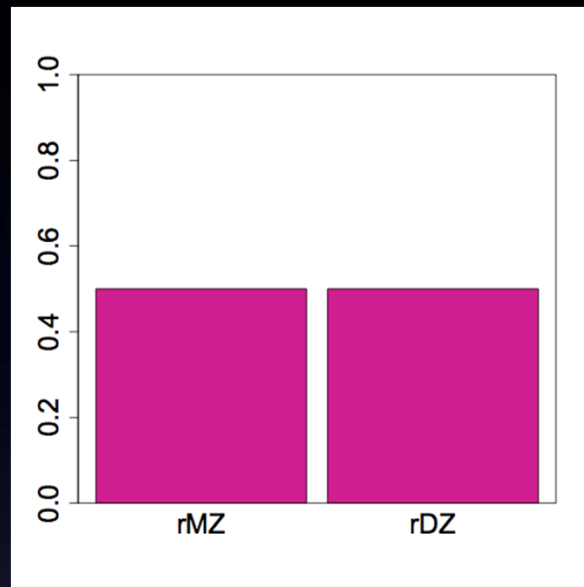
This is all great, but we still don't understand the genetic and environmental contributions to BMI

Attempts of Twin Correlation

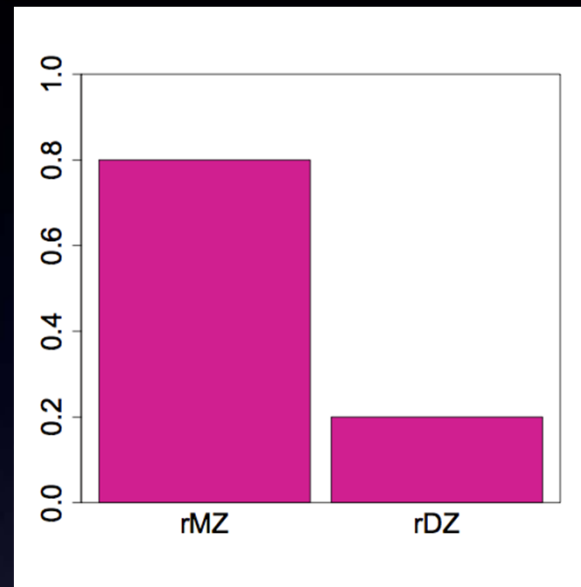


$r_{MZ} = 2r_{DZ}$
Additive

twins on average share 50% of additive effects

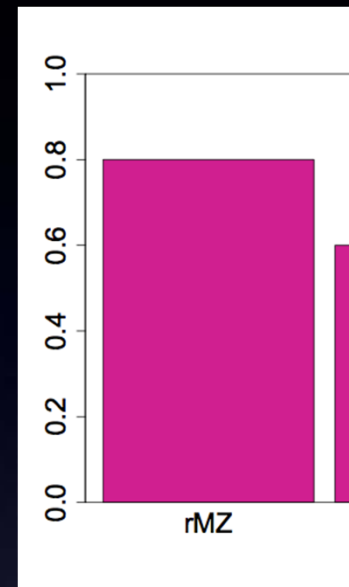


$r_{MZ} = r_{DZ}$
Shared Environment



$r_{MZ} > 2r_{DZ}$
Additive & Dominance

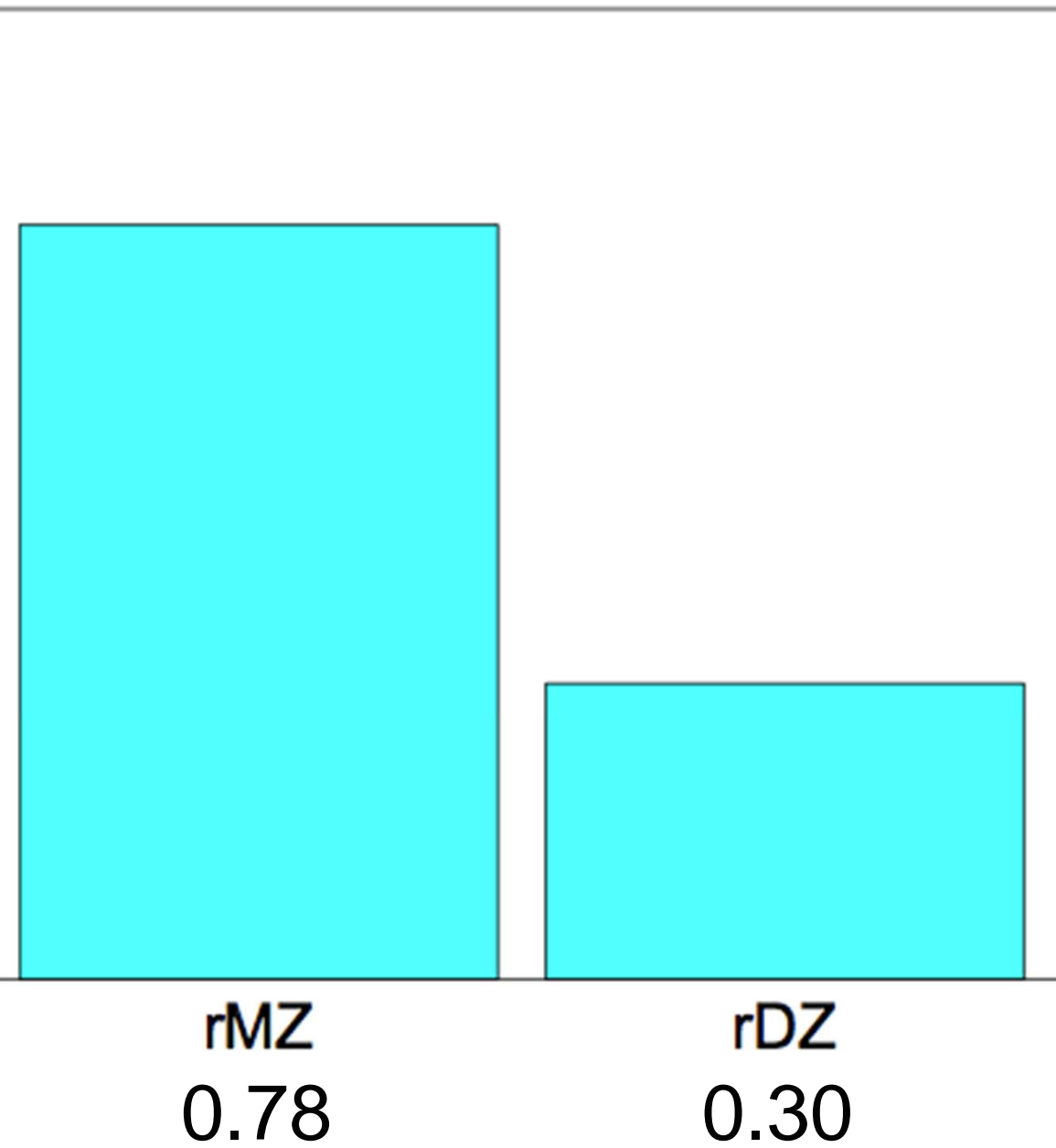
DZ twins on average share 25% of dominance



???
Additive
"Shared Environment"

$A = 2(r_{MZ} - r_{DZ})$
 $C = 2r_{DZ}$
 $E = 1 - r_{MZ}$

DIVIT with Correlations



$$A = 2(rMZ - rDZ)$$

$$C = 2rDZ - rMZ$$

$$E = 1 - rMZ$$

$$A = 0.96$$

$$C = -0.20$$

$$E = 0.22$$

ADE or ACE

A Roadmap

Use the data to test basic assumptions in the
standard ACE (ADE) models

Saturated Model

Estimate contributions of genetic and
environmental effects on the total variance of
phenotype

ADE or ACE Models

Test ADE (ACE) submodels to identify and
report significant genetic and environmental

ADE Model

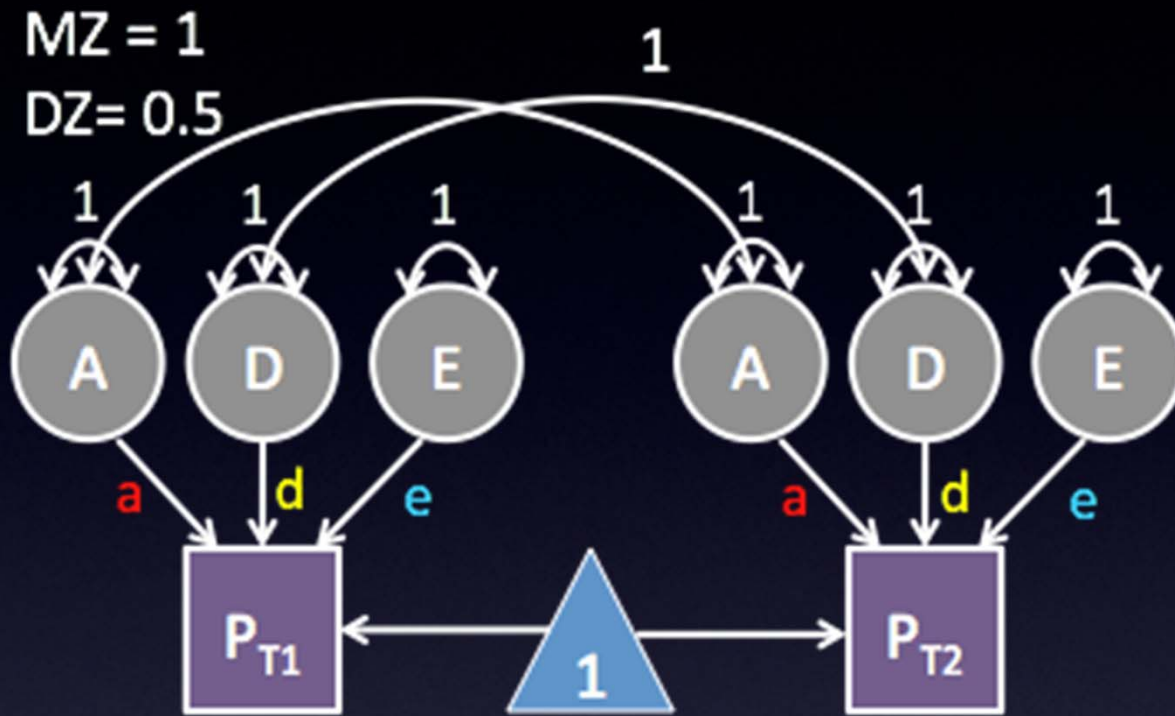
A Few Considerations

Dominance refers to non-additive genetic effects resulting from interactions between alleles at the same locus

ADE models also include effects of interactions between different loci (epistasis)

In addition to sharing on average 50% of the DNA, DZ twins also share about 25% of effects due to dominance

Path Coefficients



```
A <- mxAlgebra( expression=a %*% t(a),  
name="A" )
```

a

1 x 1 r

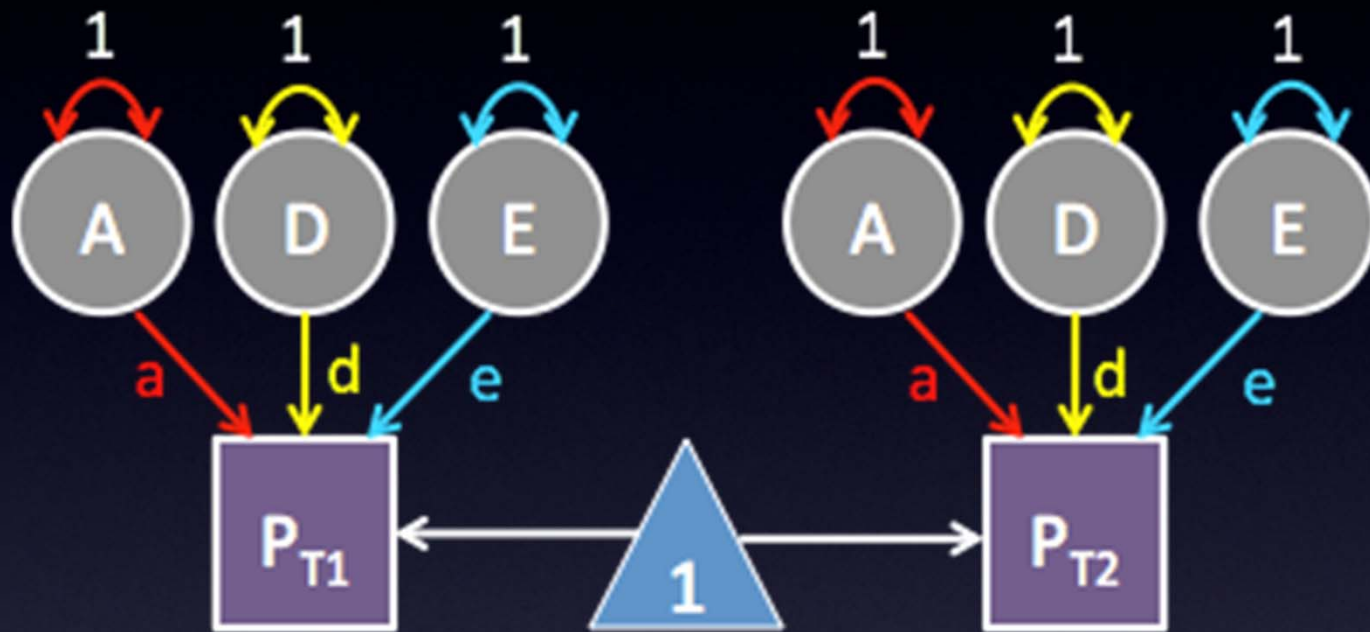
```
D <- mxAlgebra( expression=d %*% t(d),  
name="D" )
```

d

1 x 1 r

e

Variance Components



a *

1×1 matrix

d *

1×1 matrix

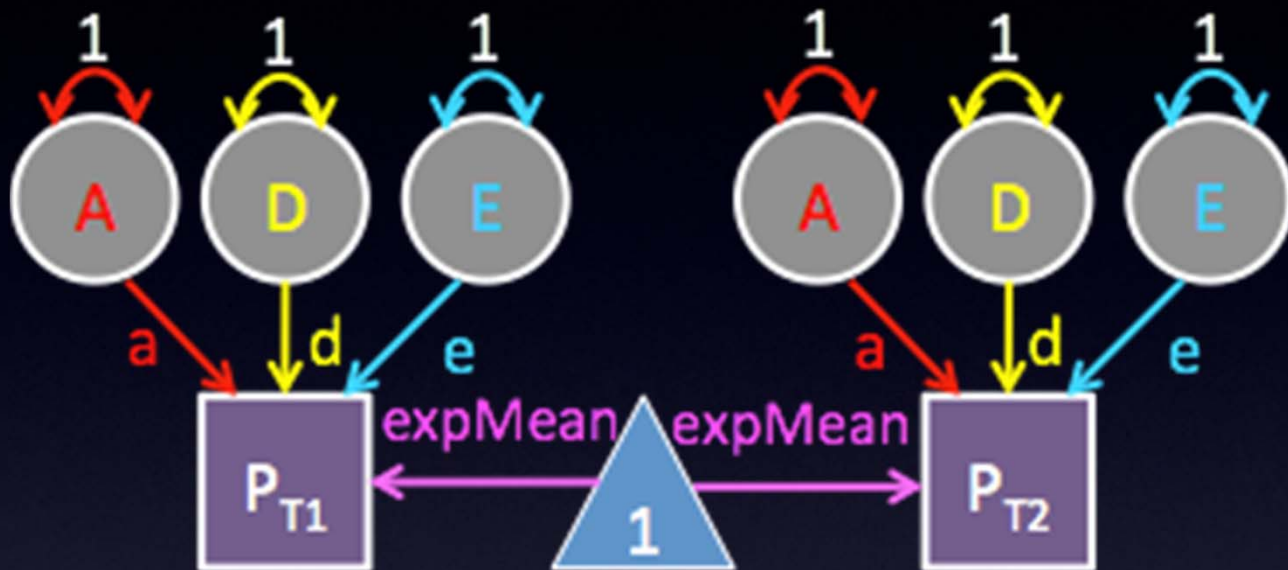
e *

```
<- mxAlgebra( expression=a %*% t(a), name="A" )
```

```
<- mxAlgebra( expression=d %*% t(d), name="D" )
```

```
<- mxAlgebra( expression=e %*% t(e), name="E" )
```

Means & Covariances



```
<- mxMatrix( type="Full", nrow=1, ncol=ntv,
             values= 20, label="mean",
             expMean"
```

expMea n	expMea n
-------------	-------------

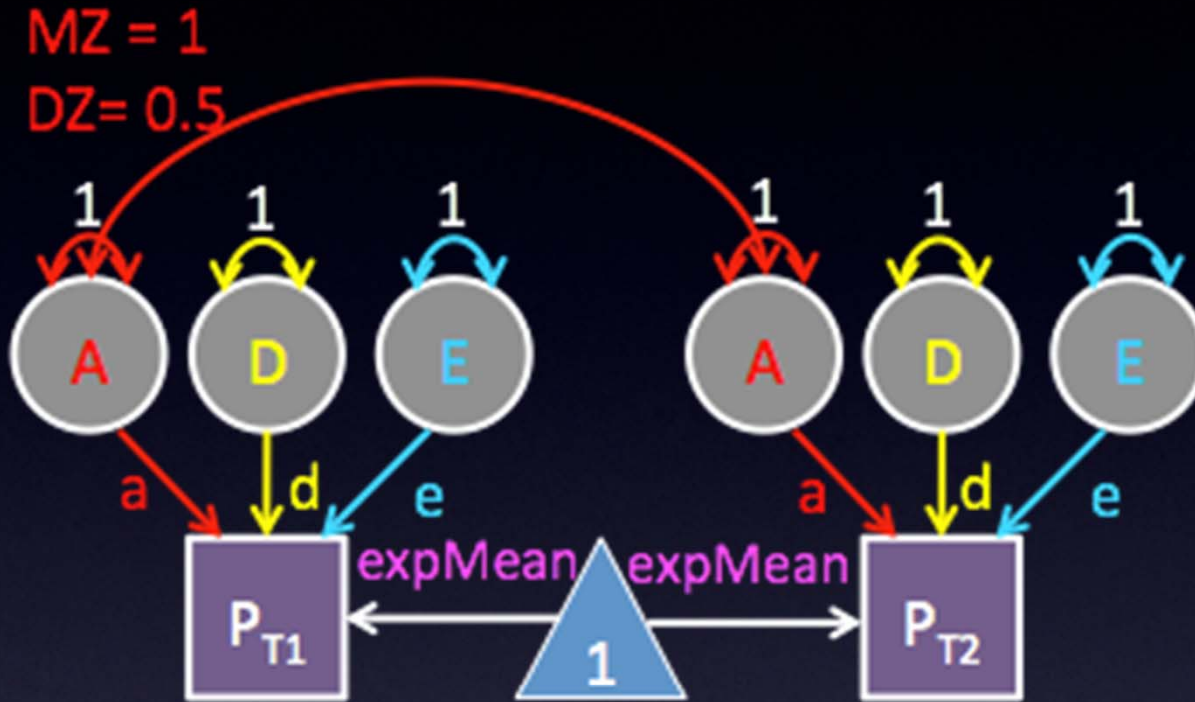
expMean= 1 x 2 m

```
<- mxAlgebra( expression= rbind( cbind(A+D+E ,
ind(A+D , A+D+E)), name="expCovMZ" )
```

```
<- mxAlgebra( expression= rbind( cbind(A+D+E,
A+0.25%x%D), cbind(0.5%x%A+0.25%x%D , A+D
```

	T1	T2
T1	A+D+E	
T2		A+D

Means & Covariances



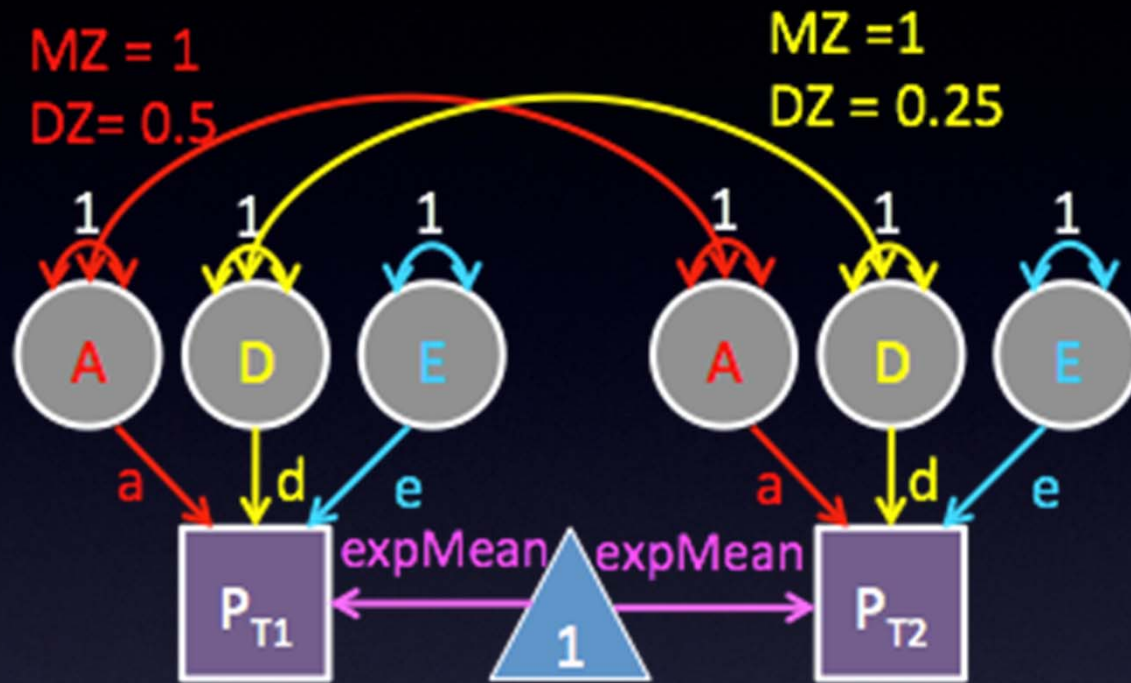
```
<- mxAlgebra( expression= rbind( cbind(A+D+E,
rbind(A+D , A+D+E)), name="expCovMZ" )
```

	T1	T2
T1	A+D+E	
T2	A	A+D

```
<- mxAlgebra( expression= rbind( cbind(A+D+E,
A+0.25%x%D), cbind(0.5%x%A+0.25%x%D,
(0.5%D+E)) , name="CovDZ")
```

	T1	T2
T1	A+D+E	
T2	0.5A	A+D

Means & Covariances



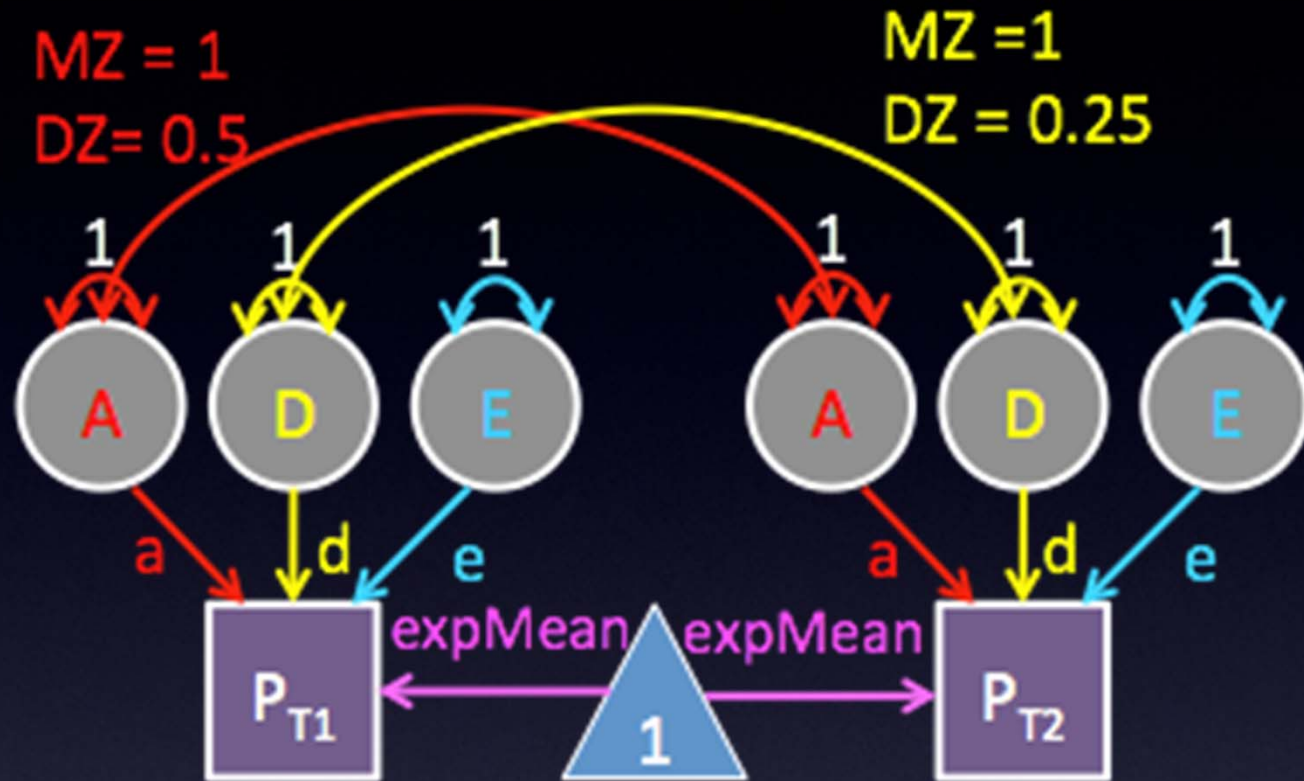
```
<- mxAlgebra( expression= rbind( cbind(A+D+E,
bind(A+D , A+D+E)), name="expCovMZ" )
```

	T1	T2
T1	A+D+E	A+
T2	A+D	A+D

```
<- mxAlgebra( expression= rbind( cbind(A+D+E,
A+0.25%x%D), cbind(0.5%x%A+0.25%x%D), (A+D
expCovDZ" )
```

	T1	T2
T1	A+D+E	0.5A+
T2	0.5A+0.25D	A+

ADL Model Decomposition



4 Parameters

Estimated

ExpMean

Variance due to A

Variance due to D

A Roadmap

Use the data to test basic assumptions in the
standard ACE (ADE) models

Saturated Model

Estimate contributions of genetic and
environmental effects on the total variance of
phenotype

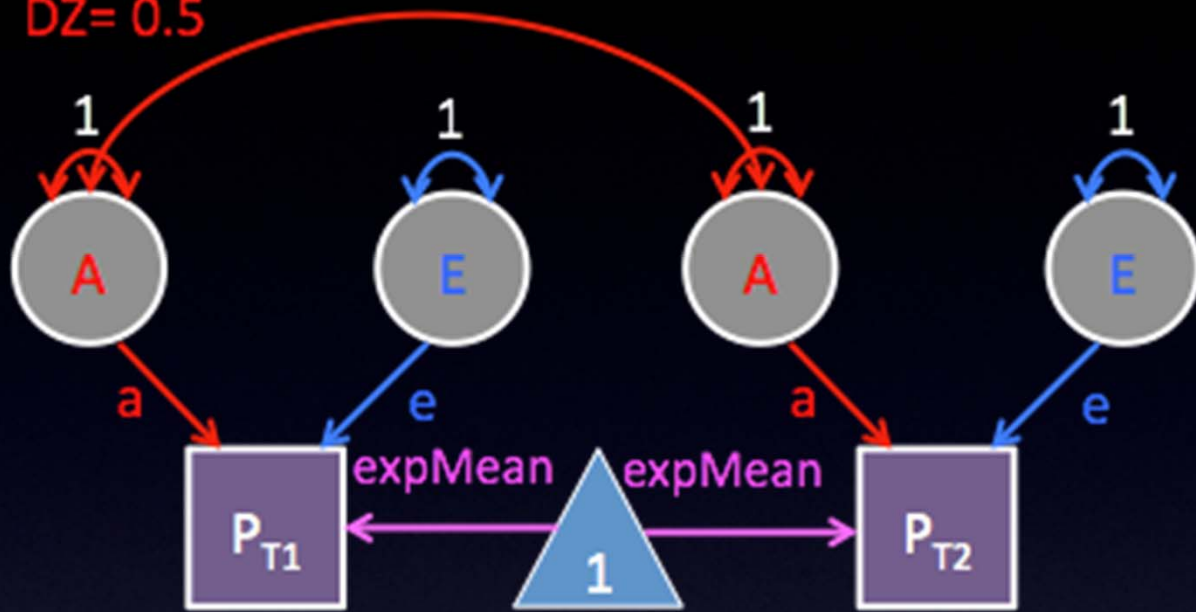
ADE or ACE Models

Test ADE (ACE) submodels to identify and
report significant genetic and environmental

Please Open twinADEcon.R

- 'Full' ADE Model
 - Comparison against Saturated
- Nested Models
 - AE Model
 - test significance of D
 - E Model vs AE Model
 - test significance of A
 - E Model vs ADE Model
 - test combined significance of A & D

MZ = 1
DZ = 0.5



Object
containing
model

Object containing
Full Model

```
AeModel <- mxModel( AdeFit, name="AE" )
```

```
AeModel <- omxSetParameters( AeModel,  
  paths="d11", free=FALSE, values=0)
```

The d
path will
estimate

Fixing the v

Goodness of Fit Stats

	ep	-2ll	df	AIC	diff - 2ll	diff df	p
saturated							
ADE							
AE							
DE	no!	no!	no!	no!	no!	no!	no!
E							

Statistics

	ep	-2ll	df	AIC	diff - 2ll	diff df	p
saturated	10	4055.93	1767	521.93	-	-	-
ADE	4	4063.45	1773	517.45	7.52	6	0.28
AE	3	4067.66	1774	519.66	4.21	1	0.06
E	2	4591.79	1775	1041.79	528.3	2	0.00

What about the magnitudes of genetic and environmental contributions to BMI?

A Roadmap

Use the data to test basic assumptions in the
standard ACE (ADE) models

Saturated Model

Estimate contributions of genetic and
environmental effects on the total variance of
phenotype

ADE or ACE Models

Test ADE (ACE) submodels to identify and
report significant genetic and environmental

Estimated Values

	a	d	e	c	a^2	d^2	e^2	c^2
ADE				-				
AE				-				
E				-				

Estimated Values

	a	d	e	c	a^2	d^2	e^2
ADE	0.57	0.54	0.41	-	0.41	0.37	0.22
AE	0.78	-	0.42	-	0.78	-	0.22
E	-	-	0.88	-	-	-	1.00

Okay, so we have a sense of which model best explains the data...or do we?

DVMT with Correlations



$$A = 2(rMZ - rDZ)$$

$$C = 2rDZ - rMZ$$

$$E = 1 - rMZ$$

ADE or ACE

Giving the ACE Model

Try

(if there is time)

Goodness-of-Fit Statistics

Model	ep	$-2ll$	df	AIC	diff - $2ll$	diff df	
saturated							
ADE							
ACE							

Goodness-of-Fit Statistics

Model	ep	-2ll	df	AIC	diff - 2ll	diff df	
saturated	10	4055.93	1767	521.93	-	-	
MADE	4	4063.45	1773	517.45	7.52	6	0
MADE	4	4067.66	1773	521.66	11.73	6	0

Estimated Values

	a	d	e	c	a^2	d^2	e^2	c^2
ADE								
AE								
ACE								
AE								
E								

Estimated Values

	a	d	e	c	a ²	d ²	e ²	c ²
ADE	0.57	0.54	0.41	-	0.41	0.37	0.22	-
AE	0.78	-	0.42	-	0.78	-	0.22	-
ACE	0.78	0.00	0.42	-	0.78	0.00	0.22	-
AE	-	-	0.56	0.68	-	-	0.41	0.59
E	-	-	0.88	-	-	-	1.00	-

Conclusions

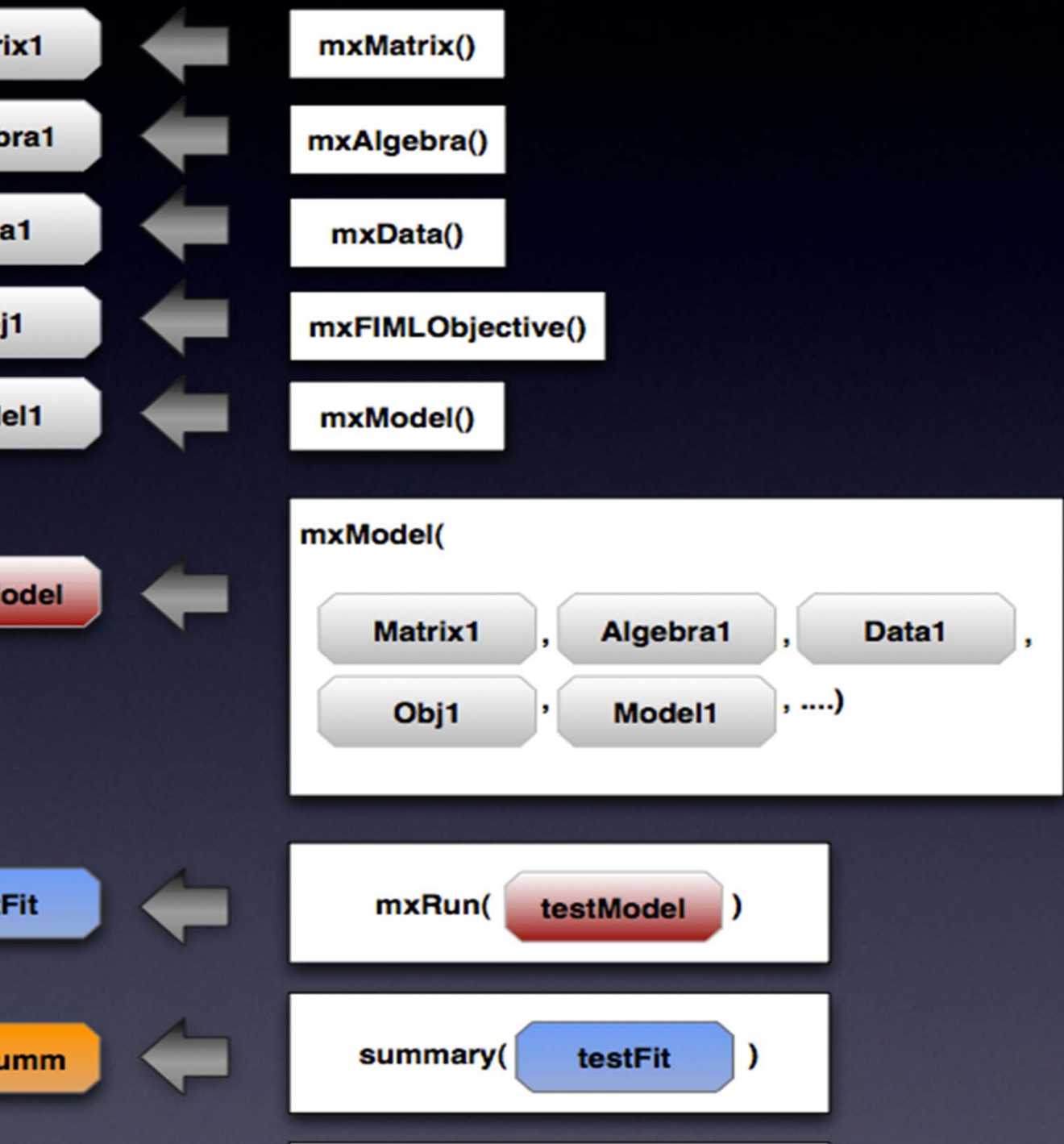
- BMI in young OZ females (age 18-30)
 - additive genetic factors: highly significant
 - dominance: borderline significant
 - specific environmental factors: significant
 - shared environmental factors: not

Potential

Complications

- Assortative Mating
- Gene-Environment Correlation
- Gene-Environment Interaction
- Sex Limitation
- Gene-Age Interaction

Processes Built the ADE Model as Coded in twinADEcon.R



Make matrices
pathA, pathD, pathE

Do Matrix Algebra w Matrices
covA, covD, covE, covMZ, covDZ

Call Data for Use in the Model
dataMZ, dataDZ

Call Data for Use in the Model
dataMZ, dataDZ

Build Model from Matrices
modelMZ, modelDZ

Build/Compile Overall Model
from Matrices/Algebras
AdeModel

Run Overall Model
AdeModel

Get Summary Information
Overall Model
AdeSumm

Generate Parameter Estimates

Thank You!