Modeling/Estimating Genetic and Environmental Components

40th International Statistical Genetics Online Workshop 2026

Hermine HM Maes et al.

Tutors: Margot Van de Weijer, Sophie Perizonius, Chandra Reynolds, Lydia Rader, Dan Gustavson, Jana Instinske, Brad Verhulst, Mike Hunter, Rob Kirkpatrick, Naomi Friedman, Tong Chen, Dinka Smajlagic, Tian Lin, Daniel Hwang, Tunde Olasege, & others

Running Day 3 Practical on YOUR COMPUTER

- You will need R (or RStudio) installed on your computer
- You will need the following R packages: OpenMx and mvtnorm
- Login to your account using
  - `ssh student@workshop.colorado.edu` (with your username in place of "`student`")
  - `cp /home/hmaes/2026/Day3/oneSACEvcrH.R .` (include the .)
  - `scp student@workshop.colorado.edu:oneSACEvcrH.R .`
- Run R script locally using R or RStudio

Schedule for Day 3 Practical

- Introduction (Recap of Video's): 20 minutes
- **First Breakout Session**: 20-25 minutes
- Brief review of testing model assumptions: 10 minutes
- Break: 10 minutes

- **Second Breakout Session**: 20-25 minutes
- Brief review of interpreting ACE model results: 10 minutes
- **Third Breakout Session**: 20-25 minutes
- Final Review: 10 minutes

Learning Objectives for Day 3 Practical

- How to test basic assumptions prior to genetic modeling
- How to fit an ACE model, interpret results and get an estimate of the twin heritability using the classical twin design
- How to setup the ACE model using the actual relatedness of twins
- What you can learn from adding genotypic data on DZ twins, with or without analyzing MZ twins

Estimating heritability

- Understanding causes of variation in phenotype & partitioning it in genetic and environmental variance components
  - Genetically informative designs -> **twin heritability**
    - infer genetic (and environmental) contributions to variance from phenotypic covariances (correlations) among family members, often twins, using expected relatedness
  - Genomically informative designs -> **SNP heritability**
    - infer genetic contributions to variance from genetic relationship matrices (correlations) based on measured genotypes of individuals, using actual relatedness

- "*Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful*"

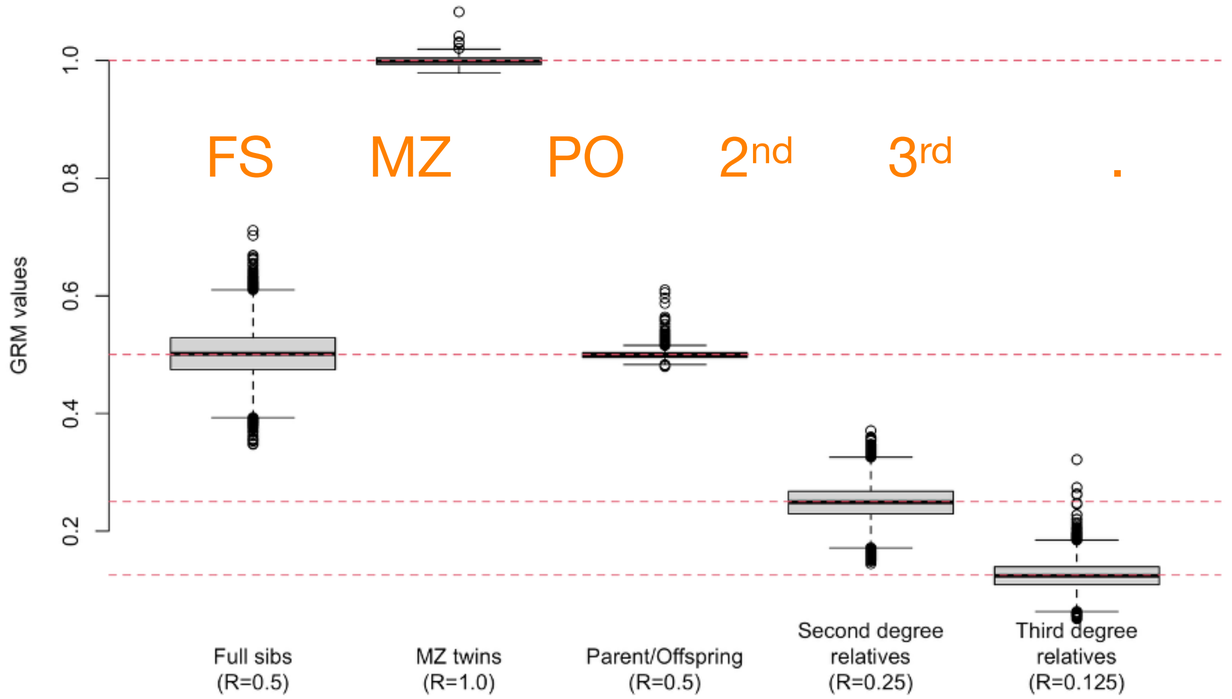
George E P Box and Norman R Draper. 1986. Empirical Model-Building and Response Surface. John Wiley & Sons, Inc., New York, NY, USA

Coefficient of Genetic Relationship > GRMs

How to estimate genetic relatedness from SNP data?

- Coefficient of genetic relationship $\mathbf{R_{jk}}$ = 2x probability that two alleles picked at random in individuals **j** and **k** are Identical-By-Descent (IBD)
- Genetic Relationship Matrix (GRM) is matrix of $\mathbf{R_{jk}}$'s and can be quantified based on expected IBD sharing (e.g., 0.5 for full sibs) or using actual SNP data

- $\hat{\pi}_{jk}$ is an estimate of $\mathbf{R_{jk}}$
- Expectation of $\hat{\pi}_{jk}$ is exactly $\mathbf{R_{jk}}$
- Observed relatedness varies by relationship



Yang, 2011

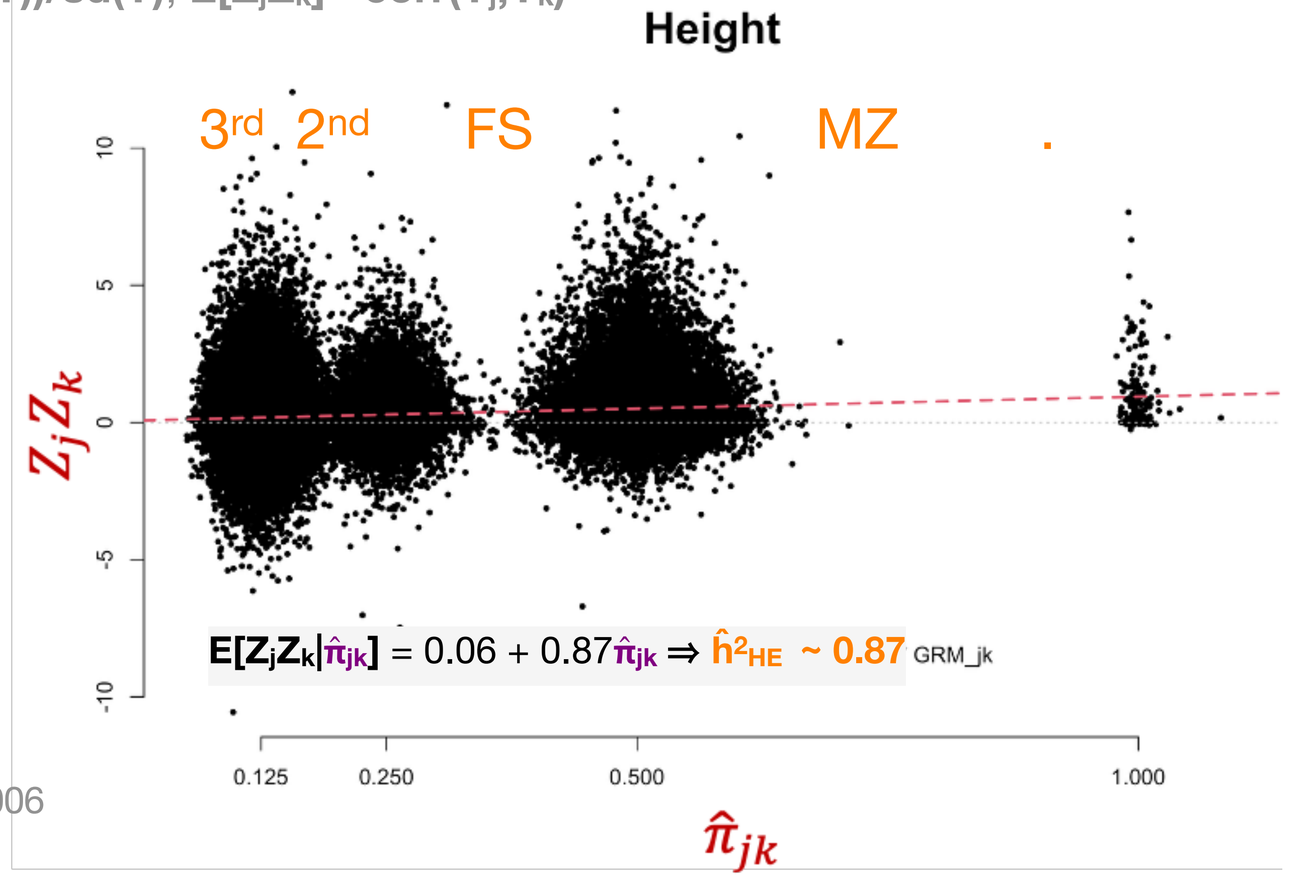
Haseman-Elston Regression

How do we use the genetic relationship coefficient to estimate SNP heritability?

- relies on association between phenotypic similarity and genetic sharing
- estimates h^2 by regressing Z_jZ_k [phenotypic similarity] onto $\hat{\pi}_{jk}$ [genetic similarity]

  where $Z_j=(Y_j-\text{mean}(Y))/\text{sd}(Y)$ & $Z_k=(Y_k-\text{mean}(Y))/\text{sd}(Y)$, $E[Z_jZ_k]=\text{corr}(Y_j,Y_k)$

- Cross-Product (CP) method:
  $Z_jZ_k = \beta_0 + \beta_1\hat{\pi}_{jk} + \varepsilon$
- Slope β_1 directly estimates heritability h^2



Haseman & Elston, 1972; Elston et al, 2000; Visscher at al, 2006

HE Regression vs Classical Twin Design

How does SNP heritability relate to Twin heritability?

- Shared goal of partitioning phenotypic variance to estimate h^2, by analyzing how phenotypic similarity between individuals relates to their genetic relatedness, with HE-regression using DNA-based relatedness rather than inferred pedigree relationships in the twin design

- **Heritability Estimation**:
  - Haseman-Elston Regression **HE** (Molecular): regresses phenotypic cross-product (CP) of siblings (or pairs of unrelated individuals) on shared proportion of alleles, estimated via Identity-By-State (IBS) or Identity-By-Descent (IBD)

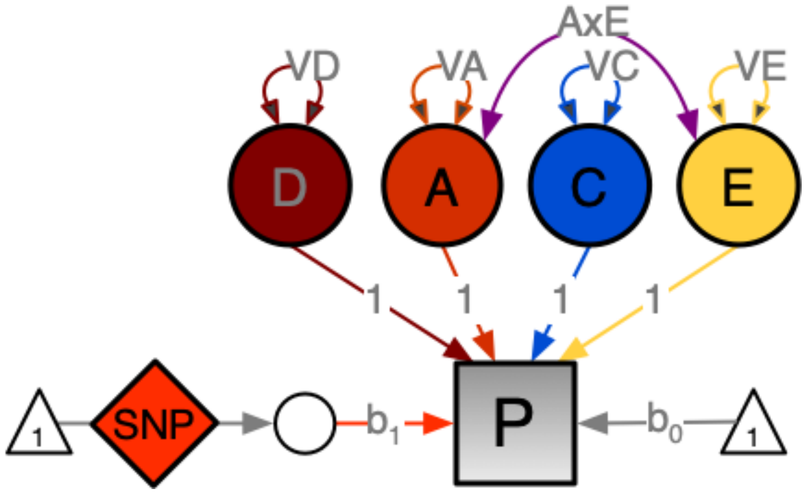
$$\mathbf{corr(Yi,Yj) = h^2 R_{jk} + Residual}$$

  - Classical Twin Design **CTD** (Traditional): compares monozygotic **MZ** twins to dizygotic **DZ** twins to partition variation in a phenotype into additive genetic **A**, common environmental **C**, and unique environmental **E** components

$$\mathbf{cov = VA \otimes rA + VC \otimes rC + VE \otimes rE}$$

Path Analysis

How to specify models using path diagrams?



- Path Diagram **Conventions**:
  - Squares or rectangles denote observed variables
  - Circles or ellipses denote latent (unmeasured) variables
  - Single-headed arrows or paths (–>): causal relationships between variables where variable at tail has influence on variable at head
  - Double-headed arrows (<–>): covariance between two variables, through common causes not in modeled; or variance of a variable
  - Triangles denote means; Diamonds denote definition variables (variables on paths)

- Path Analysis **Tracing Rules**:
  - Trace backwards, change direction at a 2-headed arrow, then trace forwards (never trace through two-headed arrows in same chain)
  - Expected covariance between two variables, or variance of a variable, is computed by multiplying together all coefficients in a chain, and summing over all possible chains

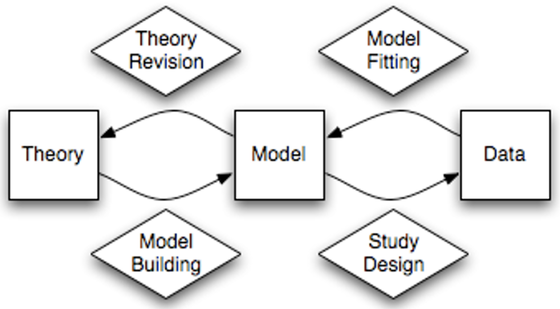
Sewall Wright, 1918; https://en.wikipedia.org/wiki/Path_analysis_(statistics)

The Classical Twin Study

How to estimate twin heritability?

- Partition phenotypic variance (V) in genetic and environmental components
  - $\mathbf{V}_{total}$ = **VG**enetic + **VE**nvironmental
  - Assumptions: additivity & independence of effects
  - Heritability ($\mathbf{h^2}$): proportion of variance due to genetic influences ($\mathbf{h^2}$ = **VG**enetic / $\mathbf{V}_{total}$)
  - Property of a group (not an individual), thus specific to that group in place & time

- Classical Twin Study/Design (CTS/CTD): monozygotic (MZ) and dizygotic (DZ) twins reared together
  - **MZ** twins share **100%** of their genes
  - **DZ** twins share on average **50%** of their genes
- Genetic factors are assumed to contribute to a phenotype when MZ twins are more similar for that phenotype than DZ twins



Eaves et al, 1978; Boomsma et al, 2002; Van Dongen et al, 2012; Polderman et al, 2015

Sources of Variance

Which factors make people different from one another?

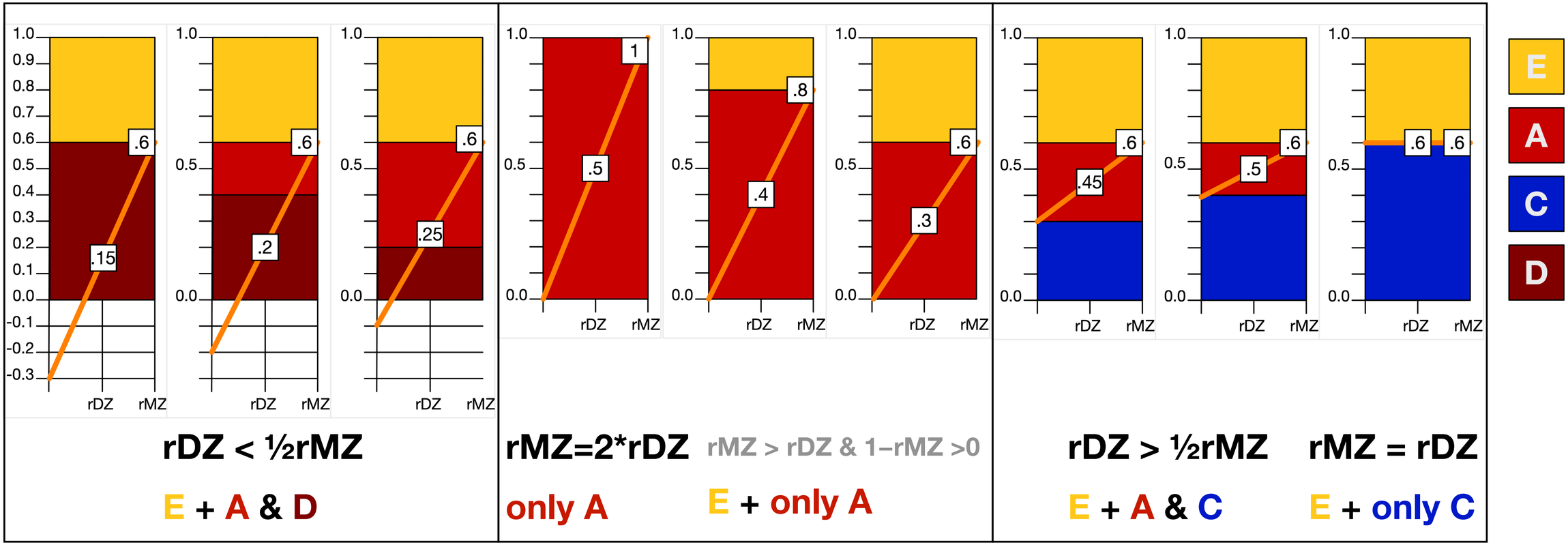
- Additive Genetic factors (**VA**, **A**, a^2): sum of all average effects of single alleles at individual loci
- Genetic Dominance (**VD**, **D**, d^2): result of interactions between alleles at same locus
- Common Environment: (**VC**, **C**, c^2): aspects of environment shared by family members, which contribute to similarity between relatives [shared]
- Unique Environment (**VE**, **E**, e^2): unique to individual, contribute to variation within family [specific, unique or within-family]

Fisher, 1918; Mather & Jinks, 1972; Martin & Eaves, 1977; Neale & Cardon, 1992

Patterns of Twin Correlations

Where does the information come from to estimate **A**, **C**, **D** & **E**?

From twin correlations to expected sources of variance



rDZ < ½rMZ

E + A & D

rMZ=2*rDZ rMZ > rDZ & 1–rMZ >0

only A **E + only A**

rDZ > ½rMZ **rMZ = rDZ**

E + A & C **E + only C**

rMZ: monozygotic twin correlation; **rDZ**: dizygotic twin correlation

A: additive genetic variance; **E**: unique environmental variance: **C**: common environmental variance; **D**: dominance genetic variance

Twin Model Assumptions

What assumptions are we making?- Important Quality Control Step!

- Equal Environments of MZ and DZ pairs (EEA)
  - MZ & DZ twins equally correlated in exposure to environmental events of etiologic importance for trait
- Random Mating, No rGE Correlation, No G x E Interaction
- No Sex Limitation, No G x Age Interaction

- **Basic Data Assumptions**: MZ & DZ twins sampled from same population, therefore we expect
  - Equal means/variances in Twin 1 and Twin 2 [twin order]
  - Equal means/variances in MZ and DZ twins [zygosity]
  - Further assumptions possibly needed when testing for heterogeneity by sex or adding other relatives [full siblings etc.]

Evans et al, 1999

Saturated Models

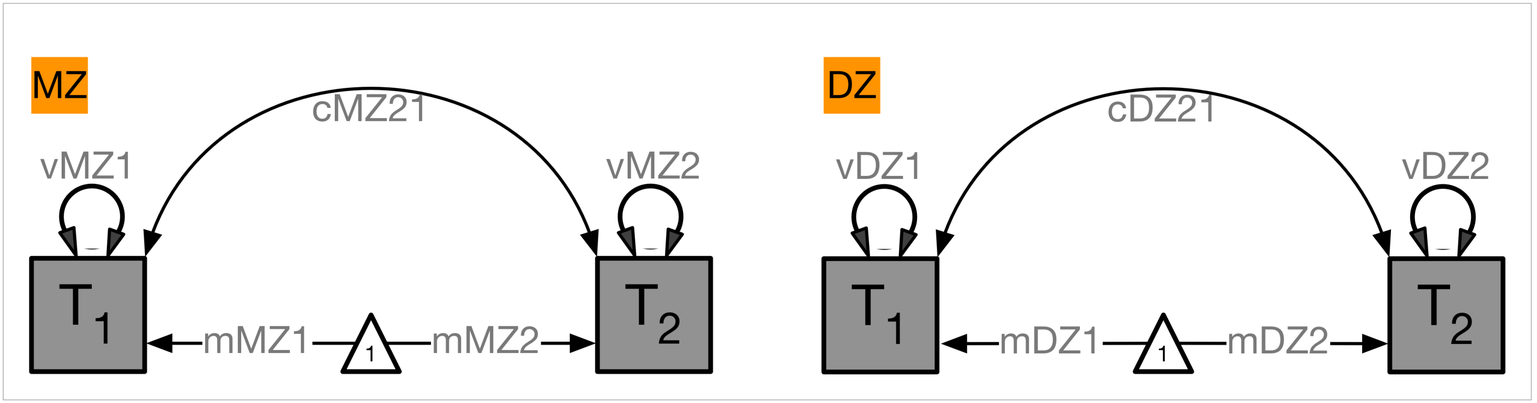
Are data assumptions valid: can we equate means/variances?

- Data: BMI in young OZ female MZ & DZ twins: observed means & variances [NH&MRC 1981 Q]

MZf twins pairs N=534		
cov	**T1**	**T2**
T1	0.73	0.59
T2	0.59	0.79
mean	21.34	21.35

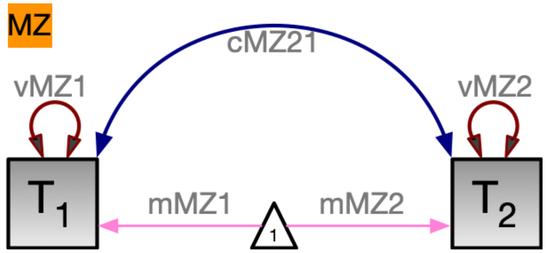
DZf twins pairs N=328		
cov	**T1**	**T2**
T1	0.77	0.24
T2	0.24	0.82
mean	21.45	21.46

- Model: as many free parameters as statistics

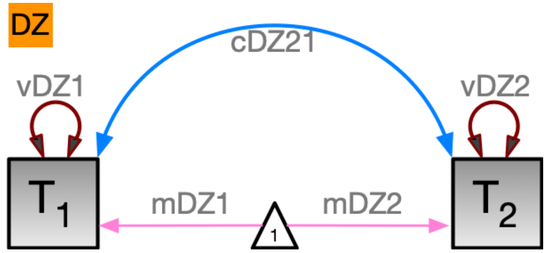


Specification of Data & Model in OpenMx

What are the key commands for OpenMx script of saturated twin model?



covMZ *2x2*	vMZ1	cMZ21
	cMZ21	vMZ2
meanMZ *1x2*	mMZ1	mMZ2



covDZ *2x2*	vDZ1	cDZ21
	cDZ21	vDZ2
meanDZ *1x2*	mDZ1	mDZ2

```
dataMZ     <- mxData( observed=mzData, type="raw" )
meanMZ     <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=svm, labels=c("mMZ1","mMZ2"),name="meanMZ" )
covMZ      <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=TRUE, values=c(svv,0,svv), labels=c("vMZ1","cMZ21","vMZ2"),
 name="covMZ" )
```

```
dataDZ     <- mxData( observed=dzData, type="raw" )
meanDZ     <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=svMe, labels=c("mDZ1","mDZ2"),name="meanDZ" )
covDZ      <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=TRUE, values=c(svv,0,svv), labels=c("vDZ1","cDZ21","vDZ2"),
 name="covDZ" )
```

Goodness of Fit Statistics - Likelihood Ratio Tests

Are data assumptions met?

		os	ep	-2ll	df	AIC	Δ-2ll	Δdf	p
	saturated	1777	10	4055.93	1767	521.93			
m1	**meanT1=meanT2**	1777	8	4056.00	1769	518.00	0.07	2	0.97
m2	**+ varT1=varT2**	1777	6	4058.94	1771	516.94	3.01	4	0.56
m3	**+ mean/varMZ=DZ**	1777	4	4063.45	1773	517.45	7.52	6	0.28

os	observed statistics
ep	estimated parameters

-2ll	-2 LogLikelihood	
df	degrees of freedom	os - ep
AIC	Akaike's Information Criterion	-2ll -2df

Δ-2ll	likelihood ratio Chi-square
Δdf	difference in df
p	probability of Chi-square

- BMI in young OZ females (age 18-30)
  - means [m1] & variances [m2] of twin 1 & 2 not significantly different in MZ & DZ pairs
  - means & variances of MZs and DZs [m3] not significantly different from one another
- Basic data assumptions of classical twin study [CTD] met

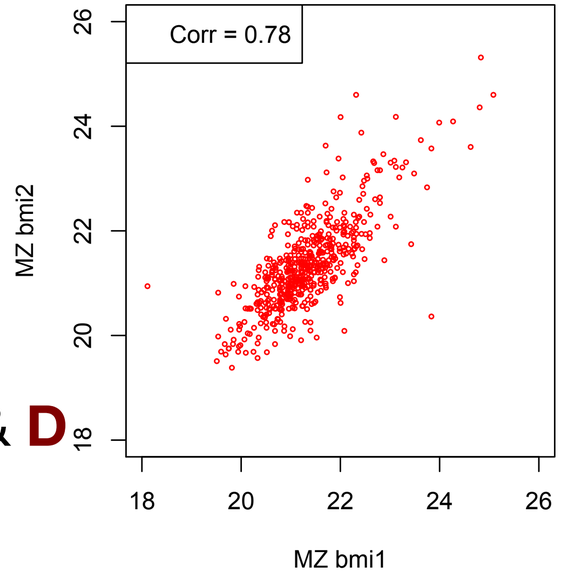
Parameter Estimates & ML Correlations

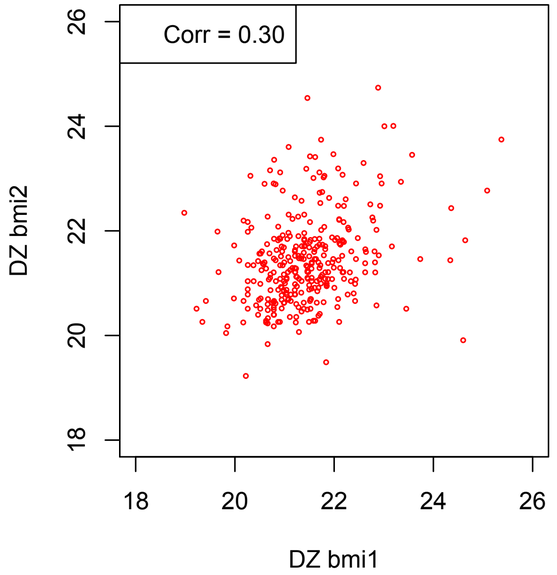
What does the pattern of MZ/DZ correlations predict about h^2?

- 4 parameters estimated in reduced model: `mZ,vZ,cMZ21 cDZ21`

MZf twins pairs N=534		
cov	**T1**	**T2**
T1	0.78	r=0.78
T2	0.61	0.78
mean	21.39	21.39

DZf twins pairs N=328		
cov	**T1**	**T2**
T1	0.78	r=0.30
T2	0.23	0.78
mean	21.39	21.39





- rMZ < 1: **E**
- rMZ > rDZ: **A** (+**D**)
- rDZ < 1/2rMZ: **A** & **D**

Practical 1

Setup data/model & Test data assumptions

- oneSACEvcrH.R
  - one: univariate script analyzing one phenotype in pairs of twins
  - S: Saturated model: freely estimating all model parameters: means, variances, covariance
  - ACE: ACE model: A:additive genetic, C:common/shared environment, E:unique/specific environment
  - v: direct variance estimation
  - c: continuous data
  - r: relatedness (both expected: rel & actual: piH)

- Section 1: Load Libraries & Functions
- Section 2: Simulate Data
- Section 3: Prepare Data
- Section 4: Prepare & Run Saturated Model
  - Section 4b: Test Data Assumptions

Simulated Data

Are data assumptions valid: can we equate means/variances?

- Data: Simulated data for 1000 pairs of MZ & 1000 pairs of DZ twins

```
> head(tDat)
  fid zyg rel piH      age sex1 sex2         T1         T2
1   1   1   1   1 57.69656    1    1  0.8076685 -1.0082586
2   2   1   1   1 46.70106    1    1  1.3422981  1.9681887
3   3   1   1   1 62.84351    1    1 -0.7068754 -1.0471504
4   4   1   1   1 66.18861    0    0  0.7987861  1.2378847
5   5   1   1   1 55.21314    0    0  0.7816142  1.1039350
6   6   1   1   1 52.16464    1    1 -0.4011456  0.4619979
> head(tDat[1000:1007,])
      fid zyg rel       piH      age sex1 sex2          T1         T2
1000 1000   1 1.0 1.0000000 63.13769    0    0  0.36294721  1.1613742
1001 1001   2 0.5 0.4597417 53.50322    0    1  0.31238999  0.8749122
1002 1002   2 0.5 0.5084421 37.28565    0    0  0.54484969 -0.2001131
1003 1003   2 0.5 0.5065984 22.86921    1    0  0.28314825  0.7468084
1004 1004   2 0.5 0.5068894 25.50235    0    0 -0.44437191 -0.5376371
1005 1005   2 0.5 0.4639071 58.63402    1    1  0.09119432 -1.2091367
```

- Variables:
  - Family number: fid
  - Phenotypes: **T1 T2**
  - Age: age
  - Sex (0=male, 1=female): sex1 sex2
  - Assigned Zygosity (1=MZ, 2=DZ): zyg
  - Expected relatedness ((1=MZ, .5=DZ): rel
  - Genomic relatedness: piH

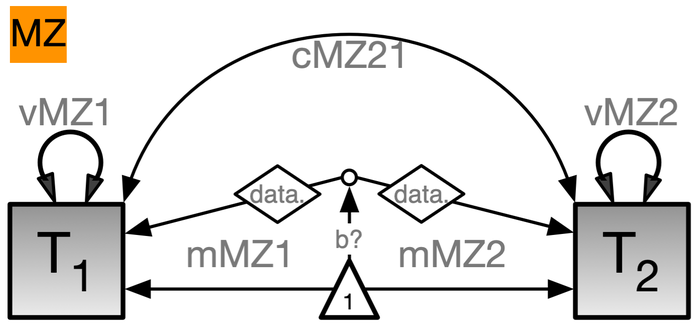
Saturated Models

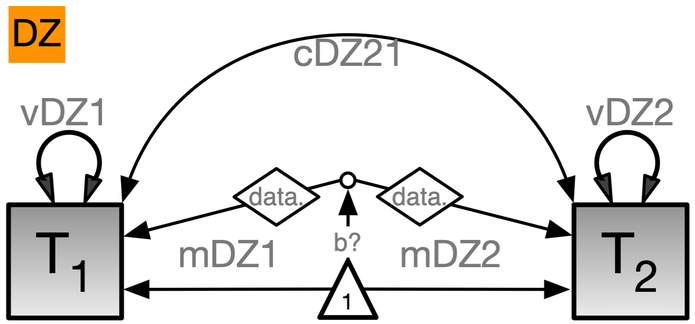
Are data assumptions valid: can we equate means/variances?

- Data: Simulated data for MZ & DZ twins: observed means & variances

MZ twins pairs N=1000		
cov	**T1**	**T2**
T1	0.97	0.67
T2	0.66	0.99
emean	*0.12	*0.15
mean	0.00	0.03

DZ twins pairs N=1000		
cov	**T1**	**T2**
T1	1.01	0.41
T2	0.40	0.93
emean	*0.01	*0.15
mean	0.01	0.03





*value of expected mean for last observation in data set

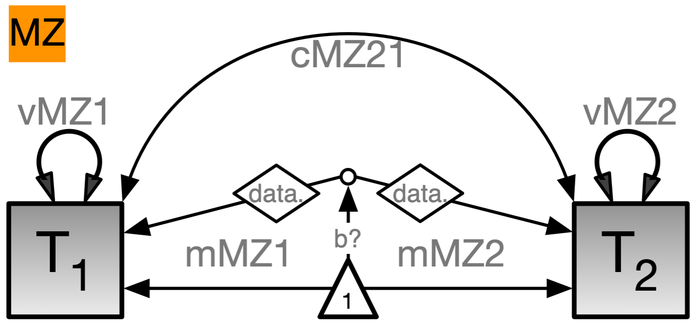
Definition Variables

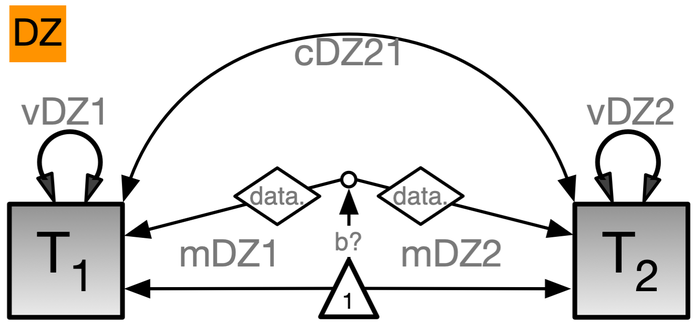
Are data assumptions valid: can we equate means/variances?

- "data.varName" in label argument of mxMatrix indicates definition variable
- matrix element updated dynamically with value from dataset when calculating row likelihood

```
> head(tDat[2000:2001,])
      fid zyg rel       piH      age sex1 sex2       T1        T2
1000 1000   1 1.0 1.0000000 63.13769    0    0 0.36294721 1.1613742
1001 1001   2 0.5 0.4597417 53.50322    0    1 0.31238999 0.8749122

betaS     <- mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, values=svBe, labels="betaS", name="bS" )
defSex    <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=FALSE, labels=c("data.sex1","data.sex2"), name="Sex" )
intercept <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=svMu, labels="interC", name="intercept" )
expMean   <- mxAlgebra( expression= intercept +bS*Sex, name="expMean" )
```



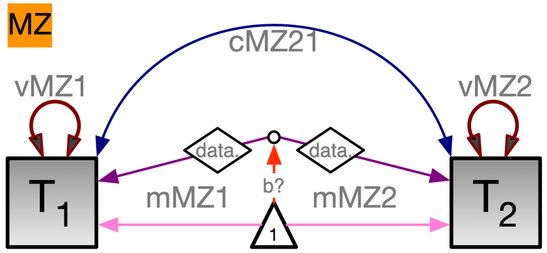


	Twin 1	Twin 2
1	μ	μ

	Twin 1	Twin 2
1	μ + b(def$_i$)	μ + b(def$_i$)

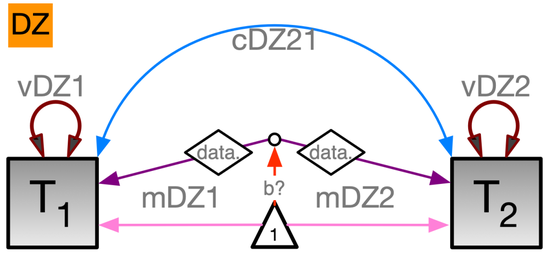
Specification of Data & Model in OpenMx

What are the key commands for OpenMx script of saturated twin model?



covMZ *2x2*	vMZ1	cMZ21
	cMZ21	vMZ2

meanMZ *1x2*	mMZ1	mMZ2



covDZ *2x2*	vDZ1	cDZ21
	cDZ21	vDZ2

meanDZ *1x2*	mDZ1	mDZ2

```
defSex     <- mxMatrix( type="Full", nrow=1, ncol=2, free=FALSE, labels=c("data.sex1","data.sex2"), name="defSex" )
defAge     <- mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels=c("data.age"), name="defAge" )
betaS      <- mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, labels="bS", name="betaS" )
betaA      <- mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, labels="bA", name="betaA" )
```

```
dataMZ     <- mxData( observed=mzData, type="raw" )
meanMZ     <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=svm, labels=c("mMZ1","mMZ2"),name="meanMZ" )
emeanMZ    <- mxAlgebra( expression= meanMZ +betaS%*%defSex +betaA%*%cbind(defAge,defAge), name="emeanMZ" )
covMZ      <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=TRUE, values=c(svv,0,svv), labels=c("vMZ1","cMZ21","vMZ2"),
 name="covMZ" )
```

```
dataDZ     <- mxData( observed=dzData, type="raw" )
meanDZ     <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=svMe, labels=c("mDZ1","mDZ2"),name="meanDZ" )
emeanDZ    <- mxAlgebra( expression= meanDZ +betaS%*%defSex +betaA%*%cbind(defAge,defAge), name="emeanDZ" )
covDZ      <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=TRUE, values=c(svv,0,svv), labels=c("vDZ1","cDZ21","vDZ2"),
 name="covDZ" )
```

Goodness of Fit Statistics - Likelihood Ratio Tests

Are data assumptions met?

		os	ep	-2ll	df	AIC	Δ-2ll	Δdf	p
	saturated	4000	12	10463.66	3988	10487.66	NA	NA	NA
m1	**meanT1=meanT2**	4000	10	10464.79	3990	10484.79	1.13	2	0.57
m2	**+ varT1=varT2**	4000	8	10466.57	3992	10482.57	2.91	4	0.57
m3	**+ mean/varMZ=DZ**	4000	6	10466.62	3994	10478.62	2.95	6	0.81

os	observed statistics
ep	estimated parameters

-2ll	-2 LogLikelihood	
df	degrees of freedom	os - ep
AIC	Akaike's Information Criterion	-2ll -2df

Δ-2ll	likelihood ratio Chi-square
Δdf	difference in df
p	probability of Chi-square

- Simulated data
  - means [m1] & variances [m2] of twin 1 & 2 not significantly different in MZ & DZ pairs
  - means & variances of MZs and DZs [m3] not significantly different from one another
- Basic data assumptions of classical twin study [CTD] met

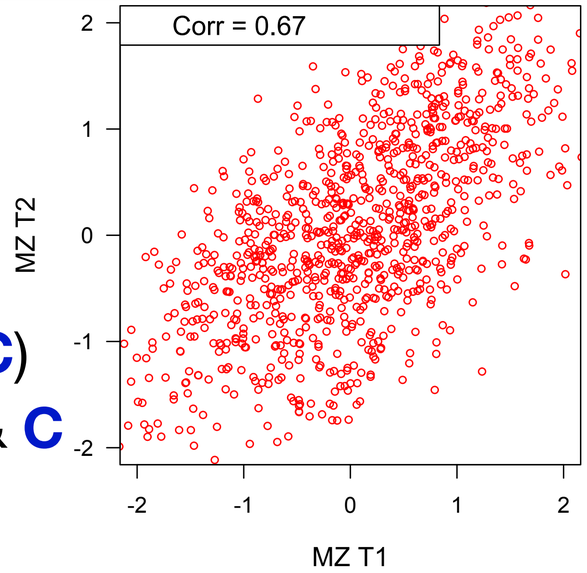
Parameter Estimates & ML Correlations

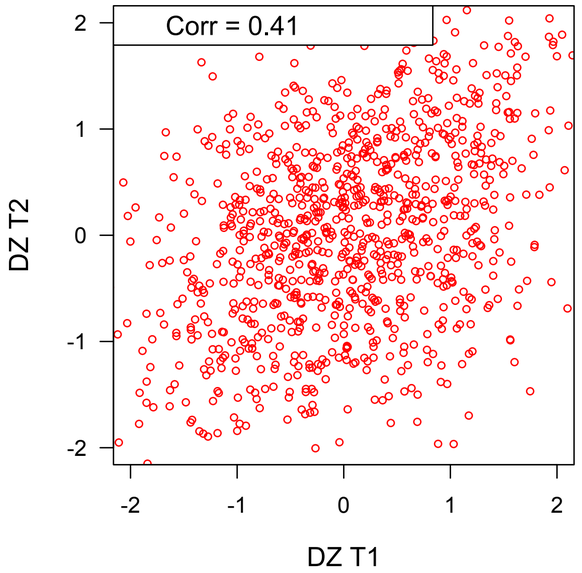
What does the pattern of MZ/DZ correlations predict about h^2?

- 4 parameters estimated in reduced model: `mZ,vZ,cMZ21 cDZ21`

MZ twins pairs N=1000		
cov	**T1**	**T2**
T1	0.97	r=0.67
T2	0.65	0.97
mean	0.02	0.02

DZ twins pairs N=1000		
cov	**T1**	**T2**
T1	0.97	r=0.41
T2	0.41	0.97
mean	0.02	0.02





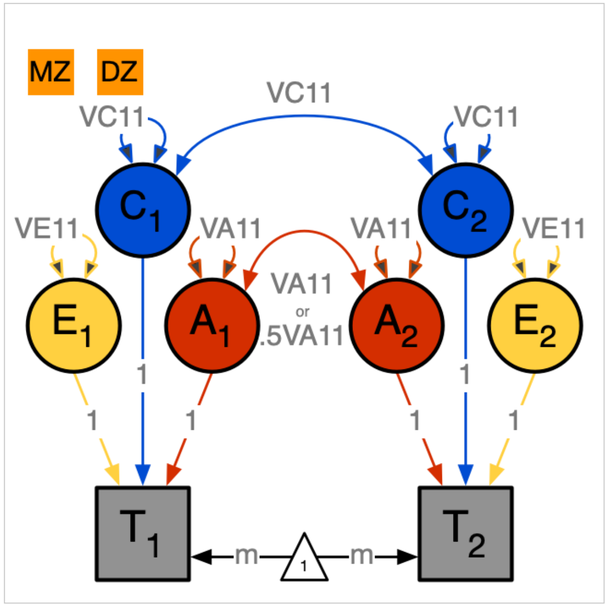
- rMZ < 1: **E**
- rMZ > rDZ: **A** (+**C**)
- rDZ > ½rMZ: **A** & **C**

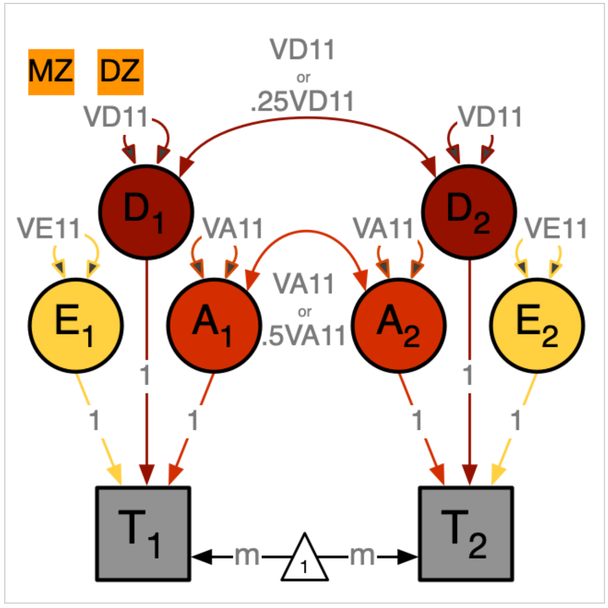
The ACE vs ADE Model

E= VE11
A= VA11
C= VC11
D= VD11

What are contributions of **A**, **C**, **D** & **E** factors to the variance?

- Classical Twin Design has three unique statistics: variance & MZ & DZ covariance, allowing estimating of three parameters: **A**, **C** & **E** or **A**, **D** & **E**
- **C** and **D** are confounded as they have opposite expectations





MZ	T1	T2
T1	A+C+E	A+C
T2	A+C	A+C+E

DZ	T1	T2
T1	A+C+E	½⊗A+C
T2	½⊗A+C	A+C+E

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

DZ	T1	T2
T1	A+D+E	½⊗A+¼⊗D
T2	½⊗A+¼⊗D	A+D+E

Univariate

- 1-rMZ: **E**
- rMZ > rDZ: **A** (+**D**)
- rMZ = 2*rDZ: only **A**
- rMZ = rDZ: only **C**
- 1/2rMZ < rDZ: **A** & **C**
- 1/2rMZ > rDZ: **A** & **D**

Path Coefficients vs Direct Variance Estimation

Why is Direct Symmetric Variance Estimation preferred?

- Path Estimation bounds variance components to be positive
  - May make incorrect inferences when comparing alternative models
- Direct (Symmetric) Variance Estimation allows variances to be estimated as negative
  - Makes correct inferences when comparing alternative models
  - Allows calculating **VD** when fitting **ACE** or calculating **VC** when fitting **ADE**

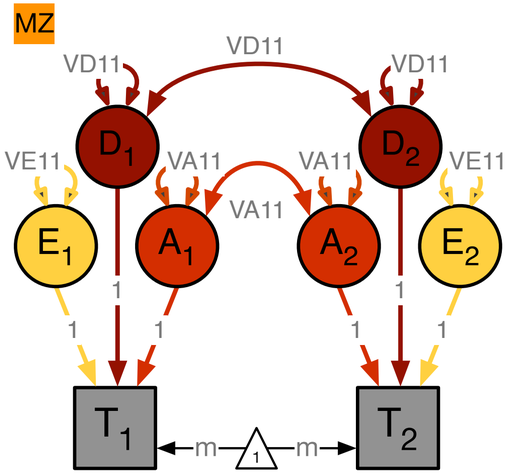
- Get **VD** when fitting **ACE** with direct variance estimation
- Three unique observed statistics
  - V = VA+VC+VE
  - cMZ = VA+VC
  - cDZ = .5VA+VC
- Three unknown parameters
  - V – cMZ = (VA+VC+VE) – (VA+VC) = VE
  - 2(cMZ – cDZ) = 2(VA+VC) – 2(.5VA+VC) = VA
  - VC = –cMZ +2cDZ, or V – VA – VE = VC
- What if negative VC?
  - VD' = –2VC
  - VA' = cMZ +2VC = VA+3VC

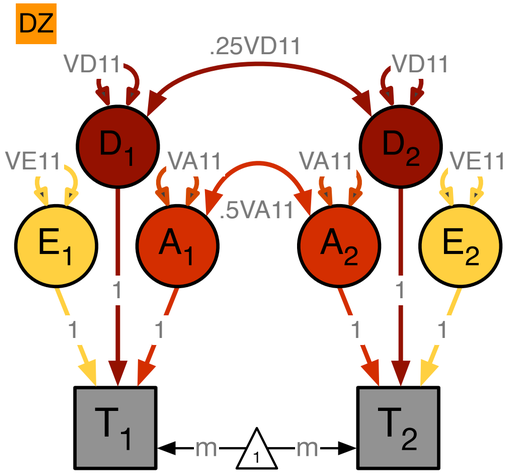
- Get **VC** when fitting **ADE** with direct variance estimation
- Three unique observed statistics
  - V = VA+VD+VE
  - cMZ = VA+VD
  - cDZ = .5VA+.25VD
- Three unknown parameters
  - V – cMZ = (VA+VD+VE) – (VA+VD) = VE
  - –(cMZ – 4cDZ) = –(VA+VD) + (2VA+VD) = VA
  - VD = 2cMZ – 4cDZ, or V – VA – VE = VD
- What if negative VD?
  - VC' = –1/2VD
  - VA' = cMZ +1/2VD = VA+3/2VD

Verhulst et al, 2019

One Trait - Univariate Twin Model Script

How to specify **ADE** model in OpenMx?





VA *1x1*	VA_{11}	***V*** *1x1*	VA+VD+VE	***expCovMZ*** *2x2*	V | cMZ cMZ | V
VC *1x1*	VD_{11}	***cMZ*** *1x1*	VA+VD	***expCovDZ*** *2x2*	V | cDZ cDZ | V
VE *1x1*	VE_{11}	***cDZ*** *1x1*	.5VA+.25VD	***meanG*** *1x2*	m | m

```
meanG    <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=svm, labels=labVars("mean",vars), name="meanG" )
covA     <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svp, label="VA11", name="VA" )
covD     <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svp, label="VD11", name="VD" )
covE     <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=sve, label="VE11", name="VE" )
covP     <- mxAlgebra( expression= VA+VD+VE, name="V" )
covMZ    <- mxAlgebra( expression= VA+VD, name="cMZ" )
covDZ    <- mxAlgebra( expression= 0.5%x%VA+ 0.25%x%VD, name="cDZ" )
ecovMZ   <- mxAlgebra( expression= rbind( cbind(V, cMZ), cbind(t(cMZ), V)), name="ecovMZ" )
ecovDZ   <- mxAlgebra( expression= rbind( cbind(V, cDZ), cbind(t(cDZ), V)), name="ecovDZ" )
```

`nv <- 1`; `svm`, `svp`, `sve`: start values for means, VA, VD, VE; `labVars`: functions in `miFunctions.R`

Interpreting Goodness of Fit - Parameter Estimates

What is the best fitting/ most parsimonious genetic model?

		os	ep	-2ll	df	AIC	Δ-2ll	Δdf	p
	saturated	1777	10	4055.93	1767	521.93			
m1	**ADE model**	1777	4	4063.45	1773	517.45	7.51	6	0.27
m2	**AE model**	1777	3	4067.66	1774	519.66	4.21	1	0.04
m3	**E model**	1777	2	4591.79	1775	1041.79	528.34	2	0.00

- ADE model fits data not significantly worse than saturated model, is most parsimonious

		variance components			standardized components		
		VA	**VD**	**VE**	**SA**	**SD**	**SE**
m1	**ADE**	0.32 0.02-0.61	0.29 0.01-0.60	0.17 0.15-0.19	0.41	0.37	0.22
m2	**AE**	0.62 0.56-0.68	-	0.17 0.17-0.19	0.78	-	0.22
m3	**E**	-	-	0.78 0.73-0.83	-	-	1.00

- Broad heritability [**VA**+**VD** / **V**] estimated at 78%, narrow heritability [**VA** / **V**] at 41%

Practical 2

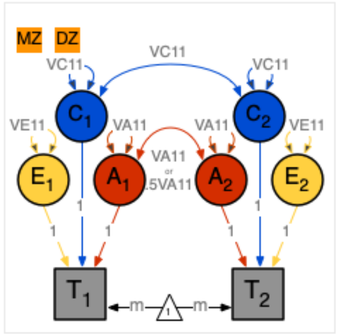
Fit multi-group ACE modeling

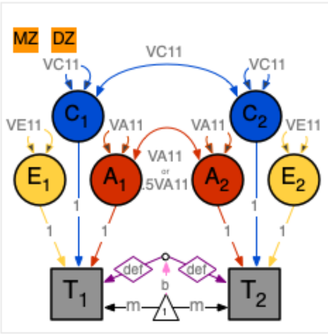
- oneSACEvcrH.R
  - one: univariate script analyzing one phenotype in pairs of twins
  - S: Saturated model: freely estimating all model parameters: means, variances, covariance
  - ACE: ACE model: A:additive genetic, C:common/shared environment, E:unique/specific environment
  - v: direct variance estimation
  - c: continuous data
  - r: relatedness (both expected: rel & actual: piH)

- Section 5: Prepare and Run ACE Model 1
  - Section 5b: Test Significance of A & C
- Section 6: Prepare and Run ACE Model 2: Alternate parameterization

Standard ACE Model

How to specify ACE model in OpenMx?






```
nv        <- 1
covA      <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svp, label="VA11", name="VA" )
covC      <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=svp, label="VC11", name="VC" )
covE      <- mxMatrix( type="Symm", nrow=nv, ncol=nv, free=TRUE, values=sve, label="VE11", name="VE" )
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" )
ecovMZ    <- mxAlgebra( expression= rbind( cbind(V, cMZ), cbind(t(cMZ), V)), name="ecovMZ" )
ecovDZ    <- mxAlgebra( expression= rbind( cbind(V, cDZ), cbind(t(cDZ), V)), name="ecovDZ" )
```

Object: covA covC covE covP covMZ covDZ

Matrix

VA + VC + VE = VA+VC+VE

VA + VC = VA+VC

.5⊗VA + VC = .5⊗VA+VC

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

	T1	T2
T1	V	cMZ
T2	cMZ'	V

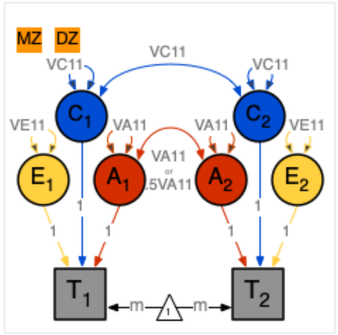
DZ	T1	T2
T1	VA+VC+VE	.5⊗VA+VC
T2	.5⊗VA+VC	VA+VC+VE

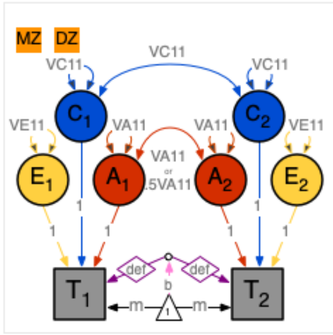
	T1	T2
T1	V	cDZ
T2	cDZ'	V

`nv <- 1`; `svm`, `svp`, `sve`: start values for means, VA, VC, VE

Output of ACE Model

How to get output from OpenMx?






```
> fitGofs(fitACE)
Mx:ACE1  os=4000  ns=2000   ep=6   co=0  df=3994  ll=10466.6152  cpu=26.6803  opt=SLSQP  ver=2.21.11  stc=0
```

```
> fitECIs(fitACE,fitACE$stVC)
   mean      bS      bA    VA11    VC11    VE11
 0.0179  0.1258 -0.0001  0.4933  0.1584  0.3229

            VA     VC     VE     SA     SC     SE      V
stVC    0.4933 0.1584 0.3229 0.5061 0.1626 0.3313 0.9747
lbound  0.3858 0.0561 0.2964 0.3972 0.0586 0.3016 0.9278
ubound  0.6066 0.2564 0.3527 0.6198 0.2597 0.3641 1.0251
```

```
> mxCompare( fitSAT, fitACE)
  base comparison ep minus2LL   df      AIC   diffLL diffdf         p
1  SAT       <NA> 12 10463.66 3988 10487.66       NA     NA        NA
2  SAT       ACE1  6 10466.62 3994 10478.62 2.951332      6 0.8149299
```

p-value > α (=0.05) thus LRT NOT significant which means that the submodel (the ACE model) is not significantly worse than full model (the Saturated model) and is more parsimonious

```
> summary(fitACE)
Summary of ACE1

free parameters:
  name matrix row col      Estimate  Std.Error A
1 mean  meanG   1   1  1.791298e-02 0.06556890
2   bS  betaS   1   1  1.257758e-01 0.03306412
3   bA  betaA   1   1 -7.507312e-05 0.00133027
4 VA11     VA   1   1  4.932850e-01 0.05603520
5 VC11     VC   1   1  1.584394e-01 0.05079627
6 VE11     VE   1   1  3.229377e-01 0.01427239

confidence intervals:
                       lbound  estimate    ubound note
ACE1.stVC[1,1] 0.38577648 0.4932850 0.6065836
ACE1.stVC[1,2] 0.05614056 0.1584394 0.2564429
ACE1.stVC[1,3] 0.29640099 0.3229377 0.3527466
ACE1.stVC[1,4] 0.39723042 0.5061087 0.6198362
ACE1.stVC[1,5] 0.05858223 0.1625583 0.2597419
ACE1.stVC[1,6] 0.30161504 0.3313330 0.3640669
ACE1.stVC[1,7] 0.92777849 0.9746621 1.0251410

Model Statistics:
               |  Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
       Model:              6                  3994               10466.62
   Saturated:             NA                    NA                     NA
Independence:             NA                    NA                     NA
Number of observations/statistics: 2000/4000

Information Criteria:
      |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
AIC:       2478.615              10478.62                10478.66
BIC:     -19891.389              10512.22                10493.16
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: 2026-06-01 01:33:50
Wall clock time: 27.65012 secs
optimizer:  SLSQP
OpenMx version number: 2.21.11
Need help?  See help(mxSummary)
```

Interpreting Goodness of Fit - Parameter Estimates

What is the best fitting/ most parsimonious genetic model?

		os	ep	-2ll	df	AIC	Δ-2ll	Δdf	p
	saturated	4000	12	10463.66	3988	10487.66			
m1	**ACE model**	4000	6	10466.62	3994	10478.62	2.95	6	0.81
m2	**AE model vs m1**	4000	5	10475.70	3995	10485.70	9.08	1	0.00
m3	**CE model vs m1**	4000	5	10552.81	3995	10562.81	86.19	1	0.00
m4	**E model vs m1**	4000	4	11250.80	3996	11258.80	784.19	2	0.00

- ACE model fits data not significantly worse than saturated model, is most parsimonious

		variance components			standardized components		
		VA	**VC**	**VE**	**SA**	**SC**	**SE**
m1	**ACE**	0.49 0.39-0.61	0.16 0.06-0.26	0.32 0.30-0.35	0.51 0.40-0.62	0.16 0.06-0.26	0.33 0.30-0.36
m2	**AE**	0.65 0.60-0.71	-	0.31 0.29-0.34	0.68 0.65-0.70	-	0.32 0.30-0.35
m3	**CE**	-	0.53 0.48-0.58	0.45 0.42-0.47	-	0.54 0.51-0.57	0.46 0.43-0.49
m4	**E**			0.97 0.93-1.02			1.00

- Broad heritability [**VA**+**VD** / **V**] estimated at 51%, narrow heritability [**VA** / **V**] at 51%

One Trait - VA ⊗ Genetic Relatedness

How to specify **ACE** model in an alternative way?

cMZ *2x2*	**VA**+VC+VE	**1**⊗**VA**+VC
	1⊗**VA**+VC	**VA**+VC+VE
cDZ *2x2*	**VA**+VC+VE	**.5**⊗**VA**+VC
	.5⊗**VA**+VC	**VA**+VC+VE
	VA+**VC**+VE	α⊗VA+**VC**
	α⊗VA+**VC**	VA+**VC**+VE
	VA+VC+**VE**	α⊗VA+VC
	α⊗VA+VC	VA+VC+**VE**

VA⊗ ***rAmz*** *2x2*	1	1
	1	1
VA⊗ ***rAdz*** *2x2*	1	.5
	.5	1
VC⊗ ***rC*** *2x2*	1	1
	1	1
VE⊗ ***rE*** *2x2*	1	0
	0	1

```
rAmz: type="Stand", nrow=2, ncol=2, free=F, values=c(1,1,1)
rAdz: type="Stand", nrow=2, ncol=2, free=F, values=c(1,.5,1)
rC: type="Unit", nrow=2, ncol=2, free=F
rE: type="Iden", nrow=2, ncol=2, free=F
ecovMZ: VA%x%rAmz +VC%x%rC +VE%x%rE
ecovDZ: VA%x%rAdz +VC%x%rC +VE%x%rE
```

Alternative

ecovMZ= VA⊗rAmz + VC⊗rC + VE⊗rE

ecovDZ= VA⊗rAdz + VC⊗rC + VE⊗rE

VA *1x1*	VA_{11}	***V*** *1x1*	VA+VC+VE	***ecovMZ*** *2x2*	VA+VC+VE	VA+VC
					VA+VC	VA+VC+VE
VC *1x1*	VC_{11}	***cMZ*** *1x1*	VA+VC	***ecovDZ*** *2x2*	VA+VC+VE	.5⊗VA+VC
					.5⊗VA+VC	VA+VC+VE
VE *1x1*	VE_{11}	***cDZ*** *1x1*	.5⊗VA+VC			

ecovMZ *2x2*			***ecovDZ*** *2x2*		
V	cMZ		V	cDZ
cMZ	V		cDZ	V

```
covA: type="Symm", nrow=1, ncol=1, free=T, values=svp, label="VA11", name="VA"
covC: type="Symm", nrow=1, ncol=1, free=T, values=svp, label="VC11", name="VC"
covE: type="Symm", nrow=1, ncol=1, free=T, values=sve, label="VE11", name="VE"
covP: VA+VC+VE, name="V"
covMZ: VA+VC, name="cMZ"
covDZ: 0.5%x%VA+VC, name="cDZ"
ecovMZ: rbind( cbind(V, cMZ), cbind(t(cMZ), V))
ecovDZ: rbind( cbind(V, cDZ), cbind(t(cDZ), V))
```

Traditional

```
ecovMZ= rbind(cbind(VA+VC+VE, VA+VC),
              cbind(VA+VC,    VA+VC+VE)
ecovDZ= rbind(cbind(VA+VC+VE, .5VA+VC),
              cbind(.5VA+VC,  VA+VC+VE)
```

One Trait - VA ⊗ Expected Genetic Relatedness

How to specify **ACE** model in an alternative way with a definition variable?

cMZ *2x2*

VA+VC+VE	1⊗VA+VC
1⊗VA+VC	VA+VC+VE

cDZ *2x2*

VA+VC+VE	.5⊗VA+VC
.5⊗VA+VC	VA+VC+VE

VA+**VC**+VE	α⊗VA+**VC**
α⊗VA+**VC**	VA+**VC**+VE

VA+VC+**VE**	α⊗VA+VC
α⊗VA+VC	VA+VC+**VE**

VA⊗ ***rAmz*** *2x2*

1	1
1	1

VA⊗ ***rAdz*** *2x2*

1	.5
.5	1

VC⊗ ***rC*** *2x2*

1	1
1	1

VE⊗ ***rE*** *2x2*

1	0
0	1

```
rAmz: type="Stand", nrow=2, ncol=2, free=F, values=c(1,1,1)
rAdz: type="Stand", nrow=2, ncol=2, free=F, values=c(1,.5,1)
rC: type="Unit", nrow=2, ncol=2, free=F
rE: type="Iden", nrow=2, ncol=2, free=F
ecovMZ: VA%x%rAmz +VC%x%rC +VE%x%rE
ecovDZ: VA%x%rAdz +VC%x%rC +VE%x%rE
```

Alternative

ecovMZ= VA⊗rAmz + VC⊗rC + VE⊗rE

ecovDZ= VA⊗rAdz + VC⊗rC + VE⊗rE

VA+VC+VE	rel⊗VA+VC
rel⊗VA+VC	VA+VC+VE

VA+**VC**+VE	α⊗VA+**VC**
α⊗VA+**VC**	VA+**VC**+VE

VA+VC+**VE**	α⊗VA+VC
α⊗VA+VC	VA+VC+**VE**

VA⊗ ***rA*** *2x2*

1	1
rel	1

VC⊗ ***rC*** *2x2*

1	1
1	1

VE⊗ ***rE*** *2x2*

1	0
0	1

```
rA: type="Stand", nrow=2, ncol=2, free=F, labels="data.rel"

rC: type="Unit", nrow=2, ncol=2, free=F
rE: type="Iden", nrow=2, ncol=2, free=F
ecovTW: VA%x%rAmz +VC%x%rC +VE%x%rE
```

rel = 1 for MZs
rel = .5 for DZs

Alternative 2

ecovTW= VA⊗rA + VC⊗rC + VE⊗rE

One Trait - VA ⊗ Actual Genetic Relatedness

How to specify ACE model in an alternative way using pihat information?

cMZ 2x2	VA+VC+VE	1⊗VA+VC
	1⊗VA+VC	VA+VC+VE
cDZ 2x2	VA+VC+VE	.5⊗VA+VC
	.5⊗VA+VC	VA+VC+VE
	VA+VC+VE	α⊗VA+VC
	α⊗VA+VC	VA+VC+VE
	VA+VC+VE	α⊗VA+VC
	α⊗VA+VC	VA+VC+VE

VA⊗ *rAmz* 2x2	1	1
	1	1
VA⊗ *rAdz* 2x2	1	.5
	.5	1
VC⊗ *rC* 2x2	1	1
	1	1
VE⊗ *rE* 2x2	1	0
	0	1

```
rAmz: type="Stand", nrow=2, ncol=2, free=F, values=c(1,1,1)
rAdz: type="Stand", nrow=2, ncol=2, free=F, values=c(1,.5,1)
rC: type="Unit", nrow=2, ncol=2, free=F
rE: type="Iden", nrow=2, ncol=2, free=F
ecovMZ: VA%x%rAmz +VC%x%rC +VE%x%rE
ecovDZ: VA%x%rAdz +VC%x%rC +VE%x%rE
```

Alternative

ecovMZ= VA⊗rAmz + VC⊗rC + VE⊗rE

ecovDZ= VA⊗rAdz + VC⊗rC + VE⊗rE

	VA+VC+VE	π̂⊗VA+VC
	π̂⊗VA+VC	VA+VC+VE
	VA+VC+VE	α⊗VA+VC
	α⊗VA+VC	VA+VC+VE
	VA+VC+VE	α⊗VA+VC
	α⊗VA+VC	VA+VC+VE

VA⊗ *rA* 2x2	1	1
	π̂	1
VC⊗ *rC* 2x2	1	1
	1	1
VE⊗ *rE* 2x2	1	0
	0	1

```
rA: type="Stand", nrow=2, ncol=2, free=F, labels="data.piH"

rC: type="Unit", nrow=2, ncol=2, free=F
rE: type="Iden", nrow=2, ncol=2, free=F
ecovTW: VA%x%rAmz +VC%x%rC +VE%x%rE
```

Alternative 3

ecovTW= VA⊗rA + VC⊗rC + VE⊗rE

One Trait - VA ⊗ Actual Genetic Relatedness

How to specify ACE model in an alternative way using pihat information?

VA+VC+VE	π̂⊗VA+VC
π̂⊗VA+VC	VA+VC+VE

VA⊗ *rA* *2x2*

1	1
π̂	1

VA+VC+VE	α⊗VA+VC
α⊗VA+VC	VA+VC+VE

VC⊗ *rC* *2x2*

1	1
1	1

VA+VC+VE	α⊗VA+VC
α⊗VA+VC	VA+VC+VE

VE⊗ *rE* *2x2*

1	0
0	1

```
rA: type="Stand", nrow=2, ncol=2, free=F, labels="data.piH"

rC: type="Unit", nrow=2, ncol=2, free=F
rE: type="Iden", nrow=2, ncol=2, free=F
ecovTW: VA%x%rAmz +VC%x%rC +VE%x%rE
```

Alternative 3

corr(Yi,Yj) = h^2 R_{jk} + Residual

ecovTW= VA⊗rA + VC⊗rC + VE⊗rE

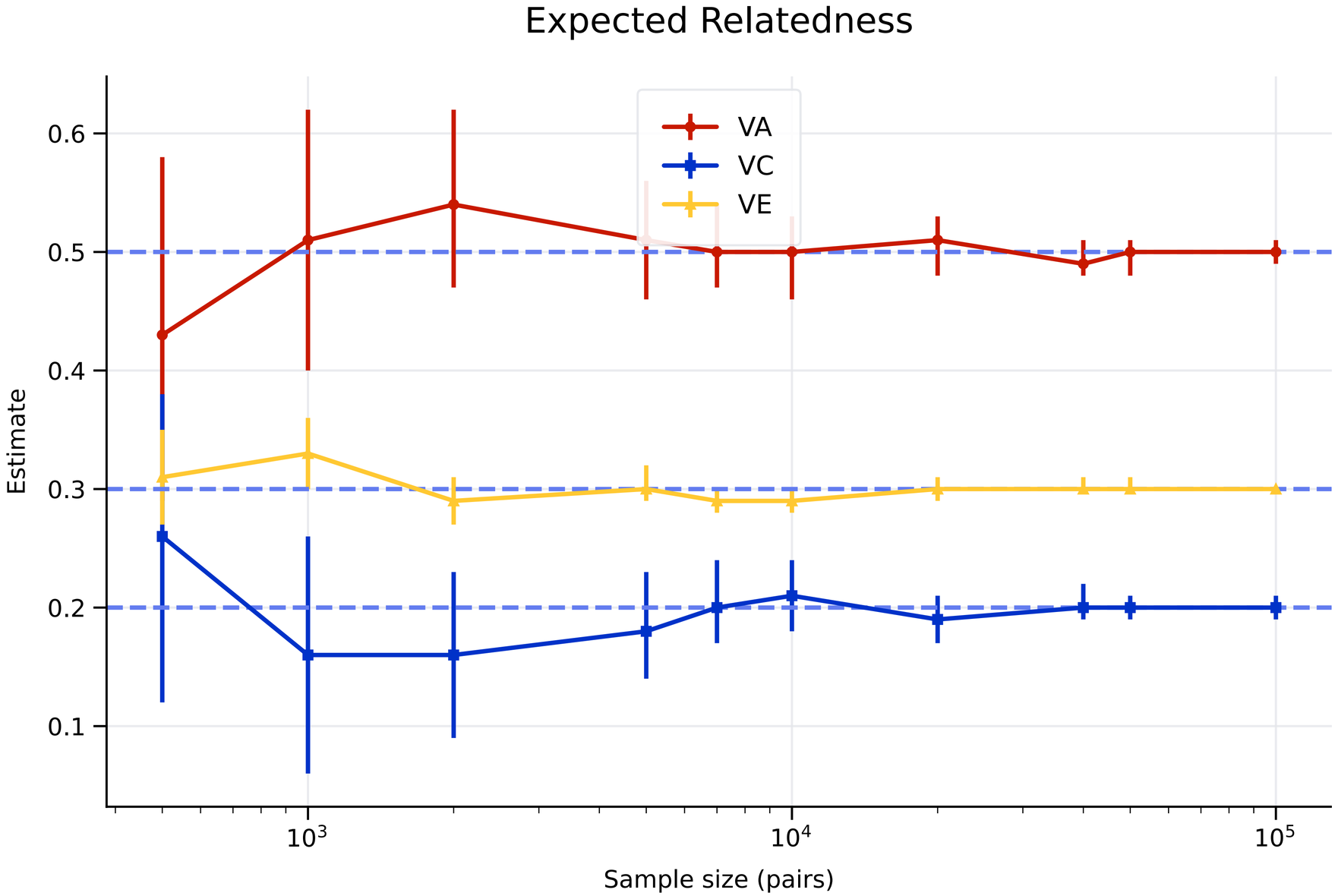
Practical 3

Fit single group ACE modeling

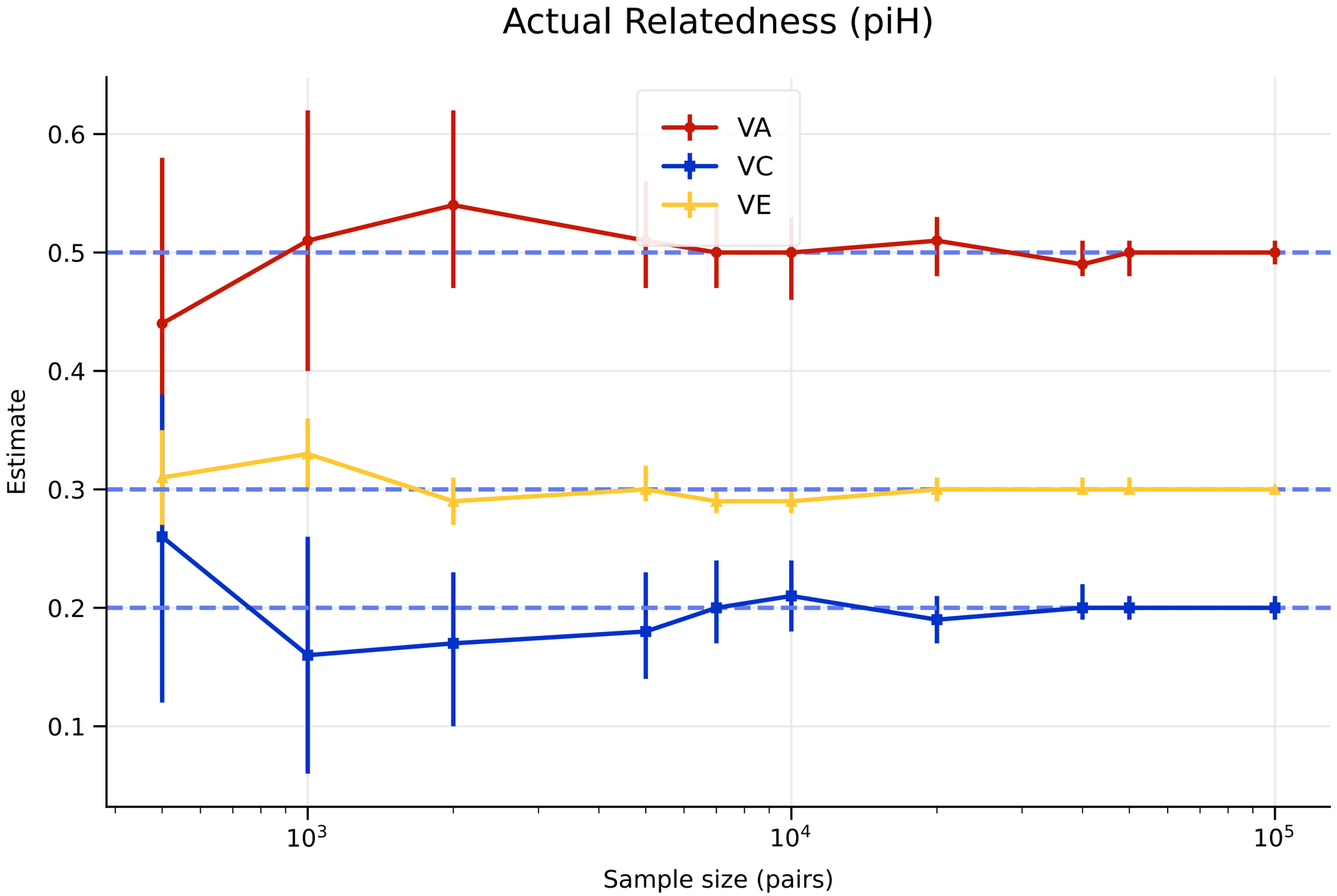
- oneSACEvcrH.R
  - one: univariate script analyzing one phenotype in pairs of twins
  - S: Saturated model: freely estimating all model parameters: means, variances, covariance
  - ACE: ACE model: A:additive genetic, C:common/shared environment, E:unique/specific environment
  - v: direct variance estimation
  - c: continuous data
  - r: relatedness (both expected: rel & actual: piH)

- Section 7: Prepare and Run ACE Model 3: Expected Relatedness as definition variable
- Section 8: Prepare and Run ACE Model 4: Actual Relatedness as definition variable
- Section 9: Prepare and Run ACE Model 3: Actual Relatedness for DZ's only
- Section 10: Summarize Results

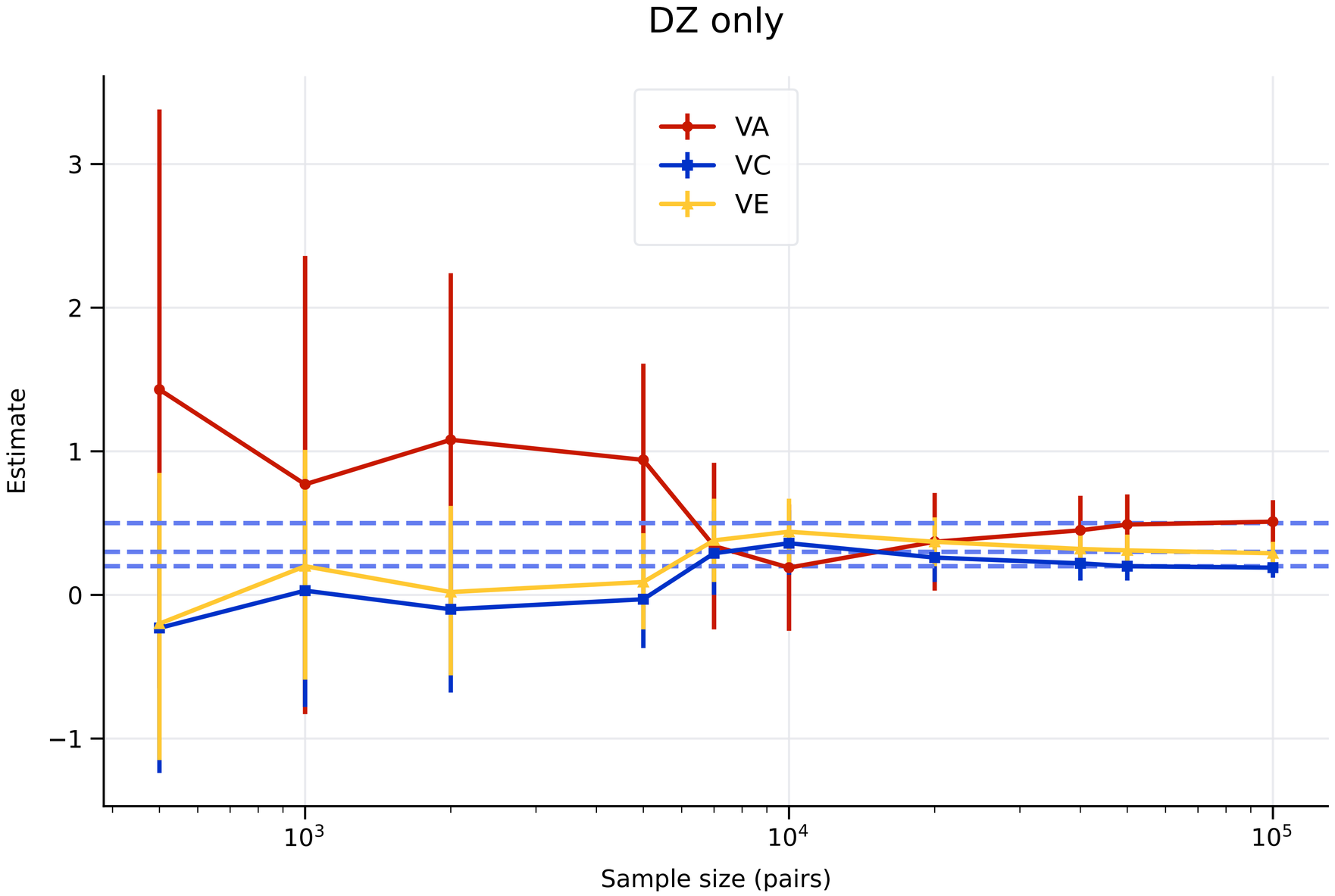
Expected Relatedness



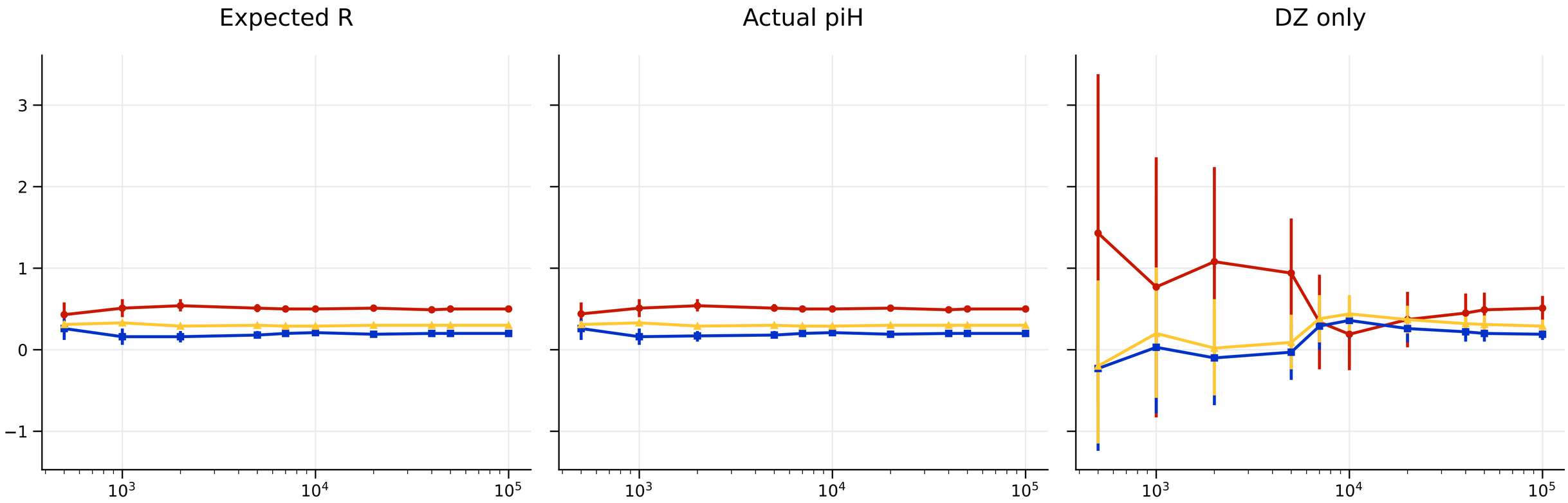
Actual Relatedness



Actual Relatedness DZs



Comparison of Three Scenarios



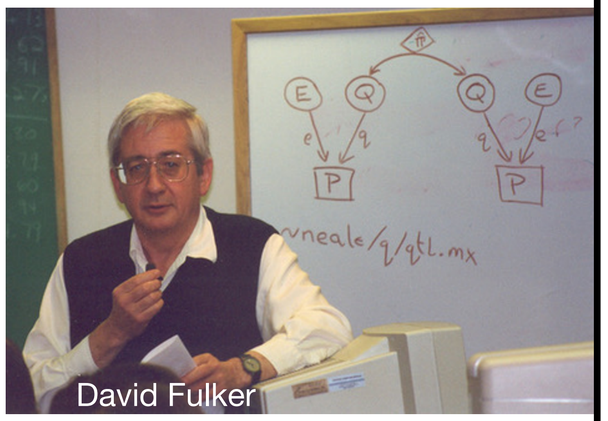
Day 3 Take Home Messages

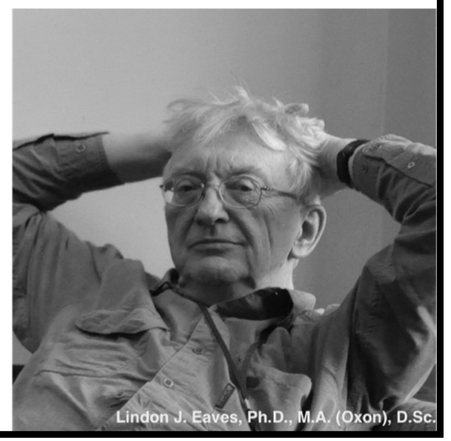
- The Classical Twin Design is a powerful design to estimate heritability, while also estimating shared and unique environmental sources of variation
- The estimate of twin heritability based on genotypic data used to calculate the actual relatedness (pihat) is similar to that using the expected relatedness when MZ data are included
- Large sample sizes are needed to get a stable estimate of the twin heritability from the actual relatedness genotyped of DZ twins (or siblings)
- Much more will be covered in the 2027 Boulder workshop

1987-2026

40th workshop

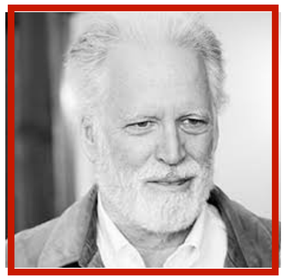
from left to right: Robert Derom†, Walter Nance†, Michael Neale, me, Joanne Meyer, Robert Vlietinck, Peter Molenaar, Andrew Heath, Karl Joreskog, Catherine Derom, Dorret Boomsma, Nick Martin, John Hewitt, Lindon Eaves†, David Fulker† [founders]





Thank you

OpenMx Development Team & International Statistical Genetics Team



Steve Boker

Mike Neale

Sarah Medland

Abdel Abdellaoui

Matt Keller

Mike Hunter

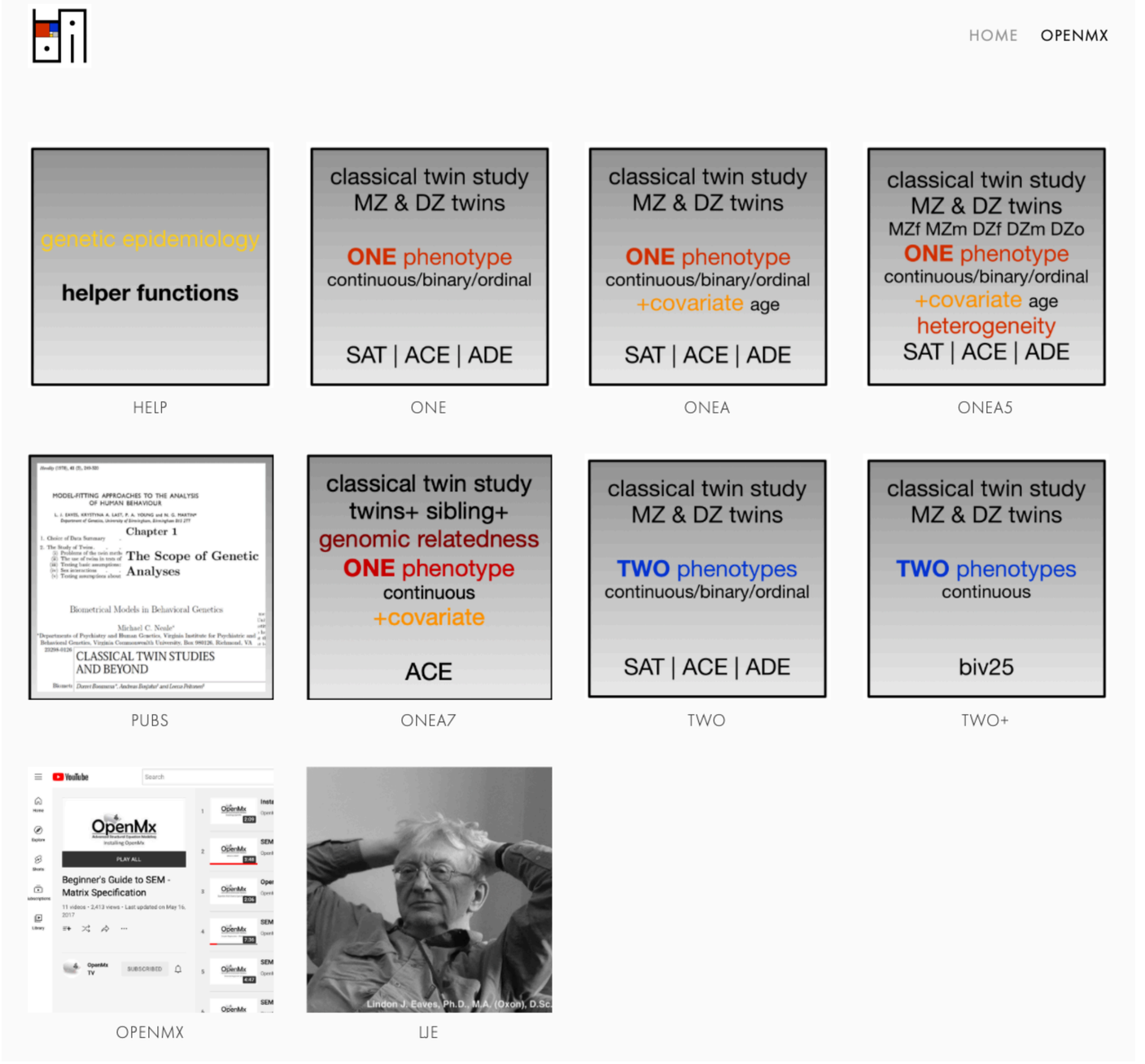
Tim Brick

Rob Kirkpatrick

John Hewitt

Jeff Lessem

hermine-maes.squarespace.com



hermine.maes@vcuhealth.org