Linear Regression Mean/Variance Estimation

- system("mkdir $HOME/hermine")
- system("cp -R /faculty/hmaes/2024/isg1.4/* $HOME/hermine")
- setwd(“~/hermine")

- OR download `oneSATc.R` [Saturated Model for one continuous phenotype] & `miFunctions.R` from https://hermine-maes.squarespace.com
- **Qualtrics**: https://qimr.az1.qualtrics.com/jfe/form/SV_0rCHbaZsRoRUgdw

ISG Workshop 2024

you can call the directory whatever you like,
it doesn’t have to be called ‘hermine’ to make it work

Regression 2 Twin Modeling

Genetic Covariance structure analysis [GCSA]
Genetically informative designs [GID]
MZ twin design [MZT]
Classical Twin Design [CTD]

Conor V Dolan, Michael C Neale, adapted by Hermine HM Maes

Outline

- Linear regression in R
- Linear regression in OpenMx
- Problem: how to infer genetic effects if you have not measured any genes?
- Genetic covariance structure analysis
- Genetically informative design - MZ twins raised together
- Observed covariance matrix vs. hypothesized covariance matrix (model)
- Representation in path diagrams vs equations
- Saturated model to test assumptions of Classical Twin Design
- Testing equality of means & variances across twin order & zygosity

Aim of Genetically Informative Designs (GID)

- **Aim:** infer genetic and environmental contributions to phenotypic variance from phenotypic covariances (correlations) among family members
  (no measured genotypes, no measured environmental variables)
- Fitting models to phenotypic data in genetically informative designs (GID) using genetic covariance structure modeling (GCSM)
- Contributions are expressed as “variance components”, so phenotypic variance is decomposed into variance components
- Start with something familiar: linear regression model (as used in GWAS)

Linear Regression Model

predict Y from X

Equation: **$Y_i = b_0 + b_1*X_i + e_i$** e.g. GWAS: $Height_i = b_0 + b_1*SNP_i + e_i$

variables:	Y_i	dependent (predicted) in participant i (i.e. Height)
	X_i	predictor in participant i (genetic variant: a SNP)
	e_i	residual in participant i
parameters:	b_0	intercept
	b_1	slope or regression coefficient

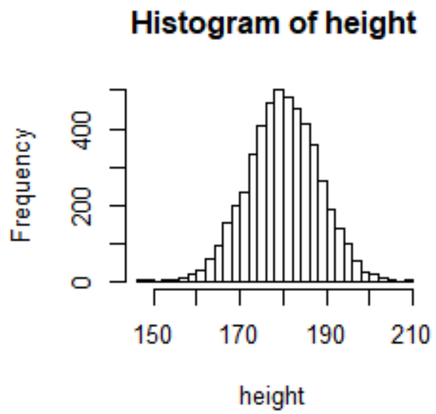
Y, X & e: variables because their value vary over persons

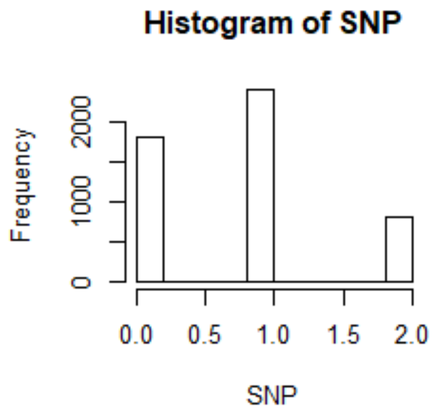
b_0 & b_1: (fixed) parameters, with unknown values (in well defined population)

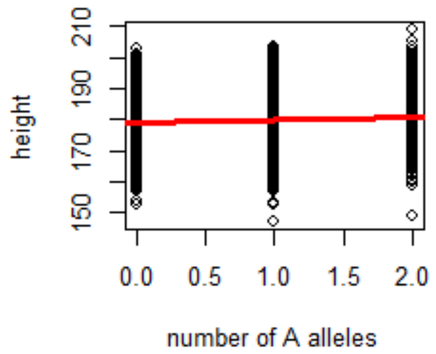
Are **X** and **Y** linearly related? **Null hypothesis (H_0): $b_1=0$**

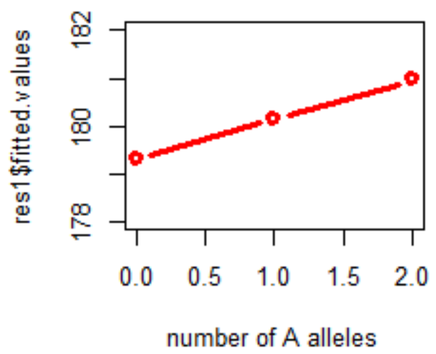
Descriptive Statistics

alleles A-a, genotypes aa, Aa/aA & AA, coded 0, 1, 2 (additive coding)









N=20,000	mean	var
height	180.03	64.01
SNP	0.80	0.48
covariance	height	SNP
height	64.010	0.212
SNP	0.212	0.480
correlation	height	SNP
height	1.000	0.038
SNP	0.038	1.000

Results of Linear Regression (in R)

SNP`[qtlA1T1]` **predicts height**`[phenoT1].` oneREGc_isg2024.R

```
lin1A       <-  lm(height~ SNP, data=dataSNP)
                    Estimate    Std. Error t value Pr(>|t|)
b0      (Intercept) 179.67245     0.08635 2080.67  < 2e-16 ***
b1      SNP           0.44226     0.08160    5.42 6.02e-08 ***
Multiple R-squared:  0.0014669, Adjusted R-squared:  0.001417
F-statistic: 29.378 on 1 and 19998 DF,  p-value: 6.0232e-08
```

Results of Linear Regression (in R)

SNP[qtlA1T1] predicts height[phenoT1]

```
lin1A       <-  lm(height~ SNP, data=dataSNP)
                    Estimate    Std. Error t value Pr(>|t|)
b0      (Intercept) 179.67245     0.08635 2080.67  < 2e-16 ***
b1      SNP           0.44226     0.08160    5.42 6.02e-08 ***
Multiple R-squared:  0.0014669, Adjusted R-squared:  0.001417
F-statistic: 29.378 on 1 and 19998 DF,  p-value: 6.0232e-08
```

- Significant association? H_0: $b_1=0$, H_1: $b_1\neq0$, $p< \alpha$ ($p=6.02e-08$, $\alpha=0.01$): reject H_0
  - Individual differences in height are linearly related to individual differences in SNP; or SNP explains variance of height or SNP is associated with height
- Linear additive model: effect of alleles A on height is additive
  - from aa (0) to Aa (1) +.44 (b_1); from aa (0) to AA (2) .44+.44 (additive: $b_1+ b_1$)

Regression in OpenMx 

oneREGc.R 1

definition variable approach. $Height_i = b0 + b1*SNP_i + e_i$

```
# Create data object
dataRaw     <- mxData( observed=dataSNP[,c("SNP1","height")], type="raw" )

# Specify SNP as a definition variable
dataSNP     <- mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels=c("data.SNP1"), name="SNP" )

# Specify model parameters
intercept   <- mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, values=0, labels="b0", name="beta0" )
regCoef     <- mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE, values=1, labels="b1", name="beta1" )
resVar      <- mxMatrix( type="Diag", nrow=1, ncol=1, free=TRUE, values=1, labels="e", name="resVar" )

# Specify expected Mean and Covariance
expMean     <- mxAlgebra( expression= beta0 + beta1 %*% SNP, name="expMean" )
expCov      <- mxAlgebra( expression= resVar, name="expCov" )
exp         <- mxExpectationNormal( covariance="expCov", means="expMean", dimnames=c("height") )
funML       <- mxFitFunctionML()

# Build and Run regression model
uniRegModel <- mxModel("Simple Regression", dataRaw, dataSNP, intercept, regCoef, resVar, expMean, expCov, exp, funML )
uniRegFit   <- mxRun(uniRegModel)
```

load simulated data

data on paths

model parameters

regression model

load all objects

Run Regression Model 

oneREGc.R 2

test significance of b1

```
# Build and Run regression model
uniRegModel <- mxModel("Simple Regression", dataRaw, dataSNP, intercept, regCoef, resVar, expMean, expCov, exp, funML )
uniRegFit   <- mxRun(uniRegModel)
summary(uniRegFit)
mxGetExpected(uniRegFit,c("mean"))

# Calculate R-squared
s2SNP       <- var(dataTSNP$qtlA1T1)   # variance of SNP
Rsquared    <- mxEval(expression= beta1^2 %*% s2SNP / (beta1^2 %*% s2SNP + resVar), uniRegFit)
print(Rsquared)

# Test significance of b1
b1SigModel  <- mxModel(uniRegFit, name="b1Sig")
b1SigModel  <- omxSetParameters(b1SigModel, label="b1", free=FALSE, values=0)
b1SigFit    <- mxRun(b1SigModel)
summary(b1SigFit)
mxCompare(uniRegFit,b1SigFit)
```

→ calculate R^2

→ drop b1 from model

→ get significance of b1

Results of Linear Regression (in R &OpenMx)

R

```
               Estimate    Std. Error t value Pr(>|t|)
b0          (Intercept) 179.67245    0.08635 2080.67  < 2e-16 ***
b1          SNP           0.44226    0.08160    5.42 6.02e-08 ***
Multiple R-squared:  0.0014669,   Adjusted R-squared:  0.001417
F-statistic: 29.378 on 1 and 19998 DF,  p-value: 6.0232e-08
```

OpenMx

```
free parameters:          name matrix row col     Estimate   Std.Error A
                       1   b0  beta0   1   1 179.67173607 0.086335752 !
                       2   b1  beta1   1   1   0.44569639 0.081569831 !
                       3    e resVar   1   1  63.91128138 0.639127709 !
Model Statistics:                      |  Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
                             Model:              3                19997             139906.88
                       Number of observations/statistics: 20000/20000
Information Criteria:        |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
                       AIC:      99940.239            139940.24              139940.24
                       BIC:     -58113.705            139956.05              139949.69
R^2= 0.0014897596
               base comparison ep  minus2LL    df       AIC    diffLL diffdf             p
1 Simple Regression       <NA>  3 139906.88 19997 139912.88        NA     NA            NA
2 Simple Regression      b1Sig  2 139936.24 19998 139940.24 29.356508      1 6.0213961e-08
```

Practical I
Linear Regression

- system("mkdir $HOME/hermine")
- system("cp -R /faculty/hmaes/2024/isg1.4/* $HOME/hermine")
- setwd(“~hermine")

- `oneREGc_isg2024.R`

- **Qualtrics**: https://qimr.az1.qualtrics.com/jfe/form/SV_0rCHbaZsRoRUgdw

Covariance Structure Model

- $Height_i = b_0 + b_1*SNP_i + e_i = 179.67 + 0.442*SNP_i + e_i$
- Increase in number of A alleles (from aa to Aa and from Aa to AA) associated with increase in height of b_1=0.44 cm. As SNP is coded 0:aa / 1:Aa,aA / 2:AA and model is linear, explained variance is called additive genetic variance.

Covariance Structure Model

- $Height_i = b_0 + b_1*SNP_i + e_i = 179.67 + 0.442*SNP_i + e_i$
- Increase in number of A alleles (from aa to Aa and from Aa to AA) associated with increase in height of b_1=0.44 cm. As SNP is coded 0:aa / 1:Aa,aA / 2:AA and model is linear, explained variance is called additive genetic variance.

- Linear regression model as a covariance structure model:
  - provides an account of the covariance (correlation) of Height and SNP (remember: correlation is expression of linear association)
  - provide a decomposition of variance of Height (remember: variance is measure of magnitude of individual differences)

Covariance & Variance Decomposition

implied by linear regression model

observed numerical cov S		
	Height	SNP
Height	s^2_H=64.01	$s_{H,SNP}$=0.212
SNP	$s_{H,SNP}$=0.212	s^2_{SNP}=0.480

cov implied by regression model		
	Height	SNP
Height	$b_1^2 * s^2_{SNP} + s^2_e$	$b_1 * s^2_{SNP}$
SNP	$b_1 * s^2_{SNP}$	s^2_{SNP}

- Regression: $Height_i = b_0 + b_1 * SNP_i + e_i = 179.67 + 0.442 * SNP_i + e_i$

Covariance & Variance Decomposition

implied by linear regression model

observed numerical cov S		
	Height	SNP
Height	s^2_H=64.01	$s_{H,SNP}$=0.212
SNP	$s_{H,SNP}$=0.212	s^2_{SNP}=0.480

cov implied by regression model		
	Height	SNP
Height	$b_1^2*s^2_{SNP}+ s^2_e$	$b_1*s^2_{SNP}$
SNP	$b_1*s^2_{SNP}$	s^2_{SNP}

- Regression: $Height_i = b_0 + b_1*SNP_i + e_i = 179.67 + 0.442*SNP_i + e_i$
- Covariance: $b_1*s^2_{SNP} = 0.442*0.480 = .212$

Covariance & Variance Decomposition

implied by linear regression model

observed numerical cov S		
	Height	SNP
Height	s^2_H=64.01	$s_{H,SNP}$=0.212
SNP	$s_{H,SNP}$=0.212	s^2_{SNP}=0.480

cov implied by regression model		
	Height	SNP
Height	$b_1^2*s^2_{SNP}+ s^2_e$	$b_1*s^2_{SNP}$
SNP	$b_1*s^2_{SNP}$	s^2_{SNP}

- Regression: $Height_i = b_0 + b_1*SNP_i + e_i = 179.67 + 0.442*SNP_i + e_i$
- Covariance: $b_1*s^2_{SNP} = 0.442*0.480 = .212$
- Decomposition: $b_1^2*s^2_{SNP} = 0.442^2*0.480 = 0.0938 + s^2_e$

Covariance & Variance Decomposition

implied by linear regression model

observed numerical cov S		
	Height	SNP
Height	s^2_H=64.01	$s_{H,SNP}$=0.212
SNP	$s_{H,SNP}$=0.212	s^2_{SNP}=0.480

cov implied by regression model		
	Height	SNP
Height	$b_1^2*s^2_{SNP}+ s^2_e$	$b_1*s^2_{SNP}$
SNP	$b_1*s^2_{SNP}$	s^2_{SNP}

- Regression: $Height_i = b_0 + b_1*SNP_i + e_i = 179.67 + 0.442*SNP_i + e_i$
- Covariance: $b_1*s^2_{SNP} = 0.442*0.480 = .212$
- Decomposition: $b_1^2*s^2_{SNP} = 0.442^2*0.480 = 0.0938 + s^2_e$
- Effect size R^2: $\{b_1^2*s^2_{SNP}\}/\{b_1^2*s^2_{SNP}+s^2_e\} = 0.0938/64.01 = .00146$ or .146%
- R^2=0.00146: proportion of variance explained by additive genetic variance

From Equation to Path Diagram

via covariance structure modeling

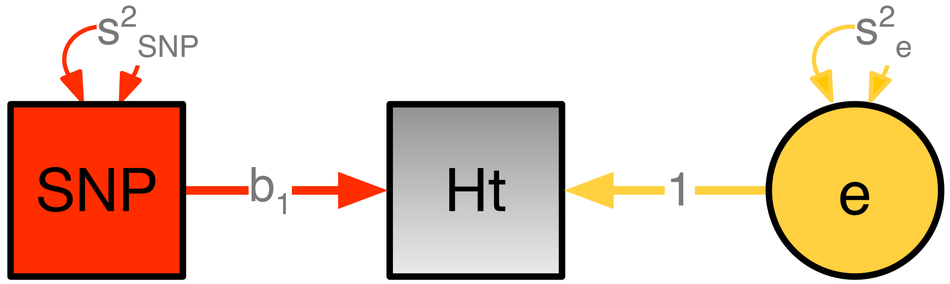
$$Height_i = b_0 + b_1 * SNP_i + e_i$$

linear regression model as equation

	Height	SNP
Height	$b_1^2 * s^2_{SNP} + s^2_e$ 64.1	$b_1 * s^2_{SNP}$ 0.212
SNP	$b_1 * s^2_{SNP}$ 0.212	s^2_{SNP} 0.480

model implied covariance structure

and numerical covariance matrix

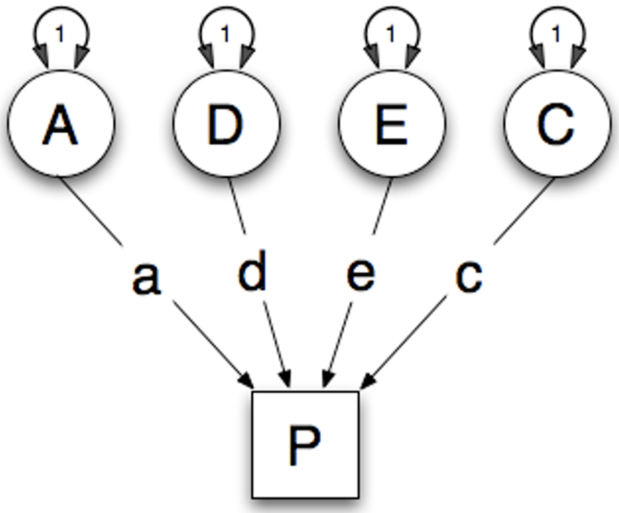


linear regression model as path diagram

Path Analysis

path diagram conventions and tracing rules

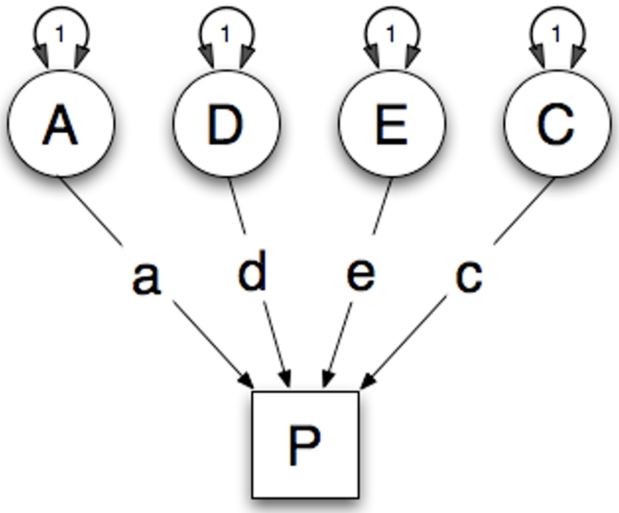
- Path Diagram **Conventions**:
  - Double-headed arrows (<–>): covariance between two variables, through common causes not in modeled; or variance of a variable
  - Circles or ellipses denote latent (unmeasured) variables
  - Upper-case letters: variables; lower-case letters: covariances/paths
  - Single-headed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



Path Analysis

path diagram conventions and tracing rules

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

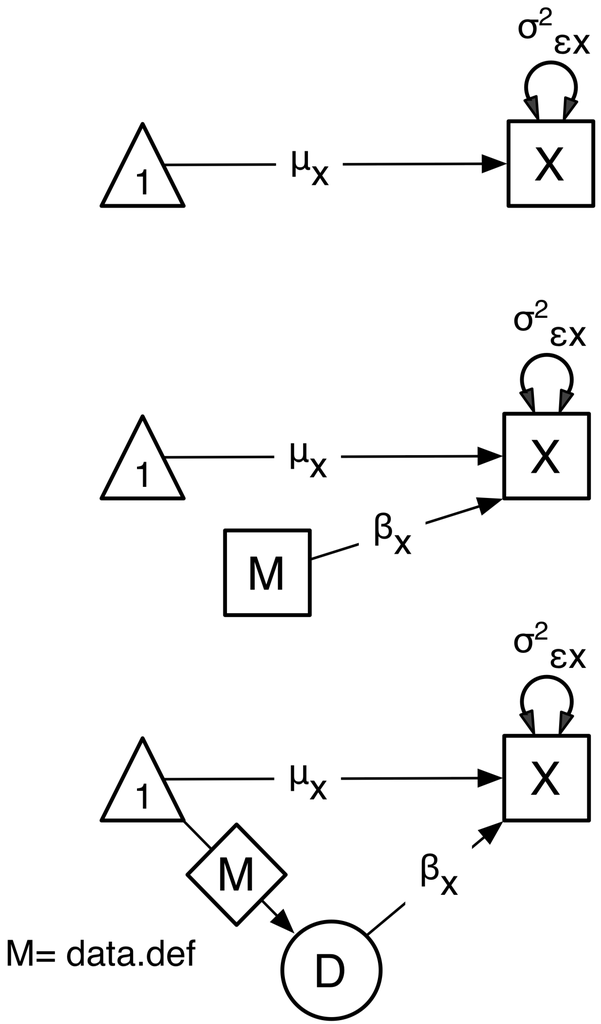


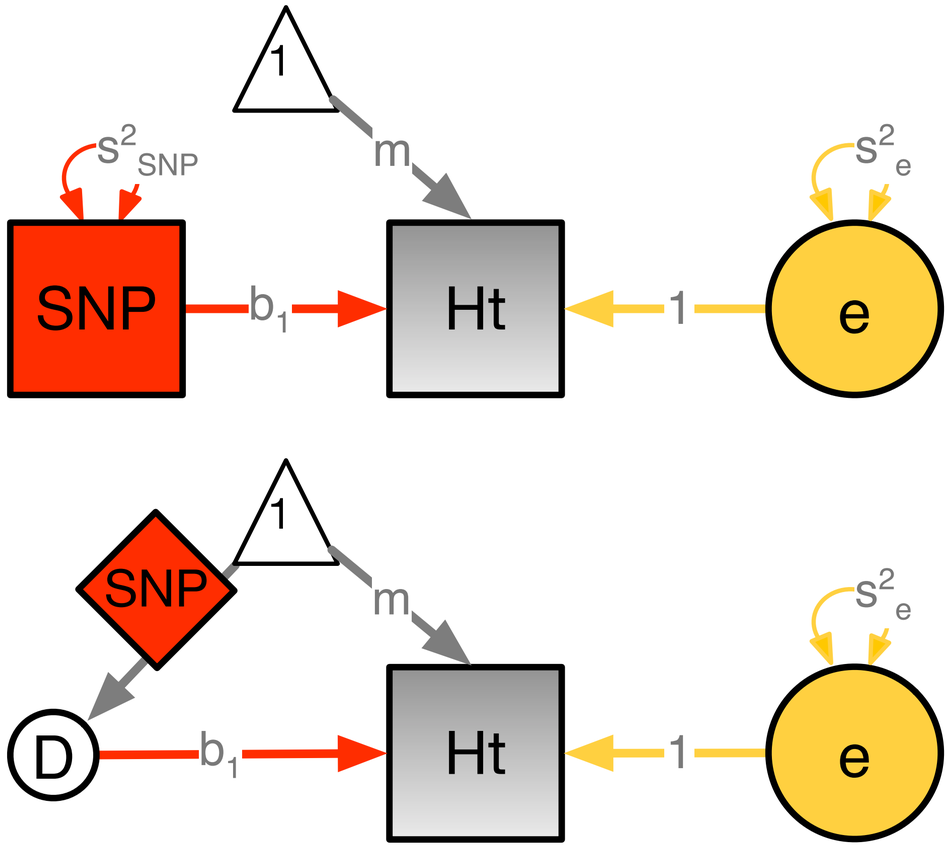
- Path Analysis **Tracing Rules**:
  - Trace backwards, change direction at a 2-headed arrow, then trace forwards (never trace through two-headed arrows in same chain)
  - Expected covariance between two variables, or variance of a variable, is computed by multiplying together all coefficients in a chain,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

Path Tracing I

definition variables

- 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

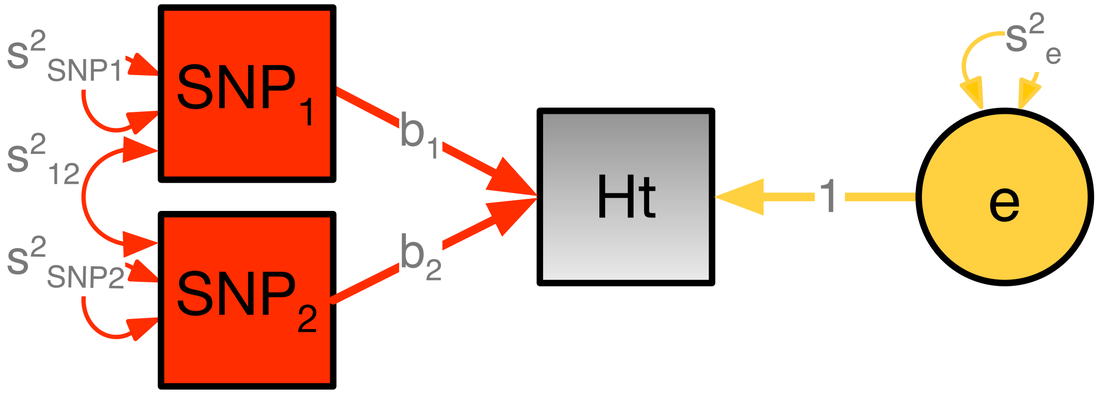




From SNPs to Additive Genetic Variance

addition of effects of multiple SNPs

- Suppose we measured all SNPs relevant to height, suppose there are just 2 SNPs
- $Height_i = b_0 + b_1*SNP1_i + b_2*SNP2_i + e_i$

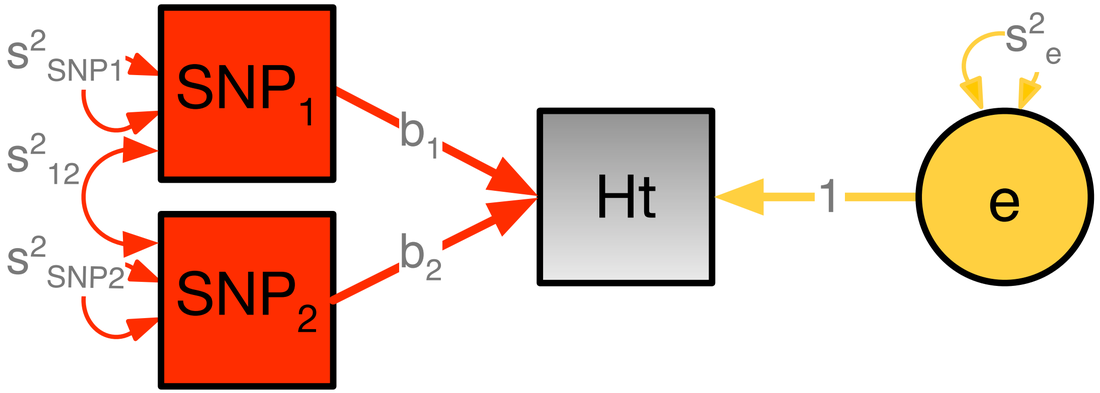


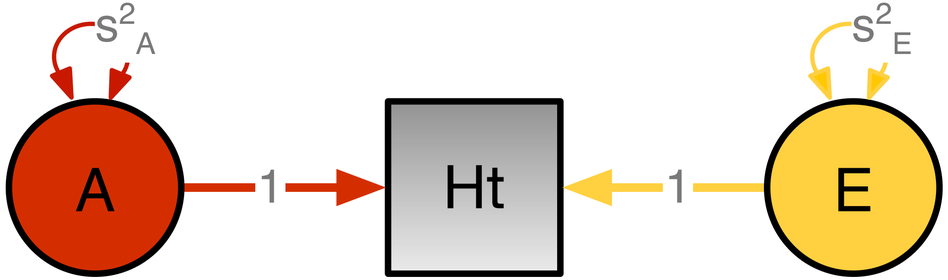
- $s^2_{Ht} = b_1^2*s^2_{SNP1}+b_2^2*s^2_{SNP2}+2*b_1*b_2*s_{12} +s^2_E$

From SNPs to Additive Genetic Variance

addition of effects of multiple SNPs

- Suppose we measured all SNPs relevant to height, suppose there are just 2 SNPs
- $Height_i = b_0 + b_1*SNP1_i + b_2*SNP2_i + e_i$ $A_i = b_1*SNP1_i + b_2*SNP2_i$





- $s^2_{Ht} = b_1^{2*}s^2_{SNP1}+b_2^{2*}s^2_{SNP2}+2*b_1*b_2*s_{12} +s^2_E$ $s^2_{Ht} = s^2_A + s^2_E$

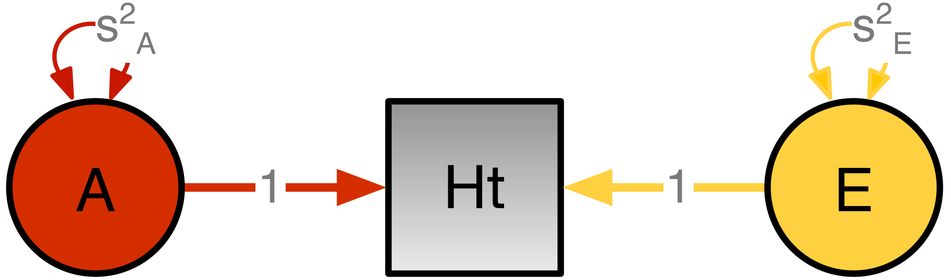
s^2_A: Additive genetic variance

s^2_E : Environmental variance

From 1 to multiple SNPs

or genetically informative design

- Suppose the following, where s^2_A is attributable to *m* SNPs (*m* >1000, say)
- Eq 1: Height = $b_0 + b_1*Gene_1 + b_2*Gene_2 + ...b_M*Gene_M + E$
- Eq 2: Height = b_0 + **A** + **E**

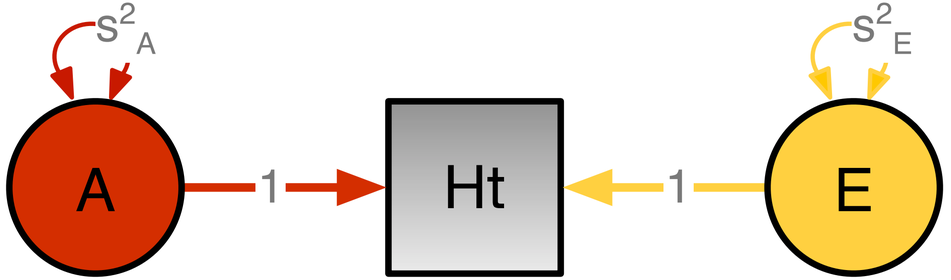


- s^2_A attributable to $Gene_1$ to $Gene_M$

From 1 to multiple SNPs

or genetically informative design

- Suppose the following, where s^2_A is attributable to *m* SNPs (*m* >1000, say)
- Eq 1: Height = $b_0 + b_1*Gene_1 + b_2*Gene_2 + ...b_M*Gene_M + E$
- Eq 2: Height = b_0 + **A** + **E**



- s^2_A attributable to $Gene_1$ to $Gene_M$
- How to estimate variance components, if we have not measured SNPs?
- Solution: Genetically Informative Design (GID)
- + Genetic covariance structure modelling (GCSM)

Genetic Covariance Structure Model [GCSM]

with Genetically Informative Design [GID]

- A model for the linear relationships among phenotypes
  - collected in a genetically informative design (GID)
  - measured in individuals in known genetic / environmental relationships
- **GID aim**: estimate genetic and environmental variance components based only on phenotype measures, no measured genes (SNPs)or environments

Polderman et al 2015 Nat Genet 47, 702–709 (2015). https://doi.org/10.1038/ng.3285

Genetic Covariance Structure Model [GCSM]

with Genetically Informative Design [GID]

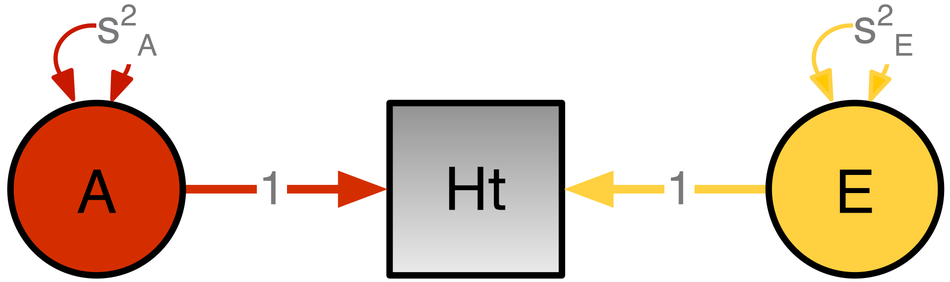
- A model for the linear relationships among phenotypes
  - collected in a genetically informative design (GID)
  - measured in individuals in known genetic / environmental relationships
- **GID aim**: estimate genetic and environmental variance components based only on phenotype measures, no measured genes (SNPs)or environments
- Most used GID: MZ and DZ twins raised together: classical twin design (CTD)
- Start with a **simpler GID**: MZ twins raised together (**MZT**)
- **Design**: collect height in a representative sample of MZ twins (i.e., representative of well-defined population)

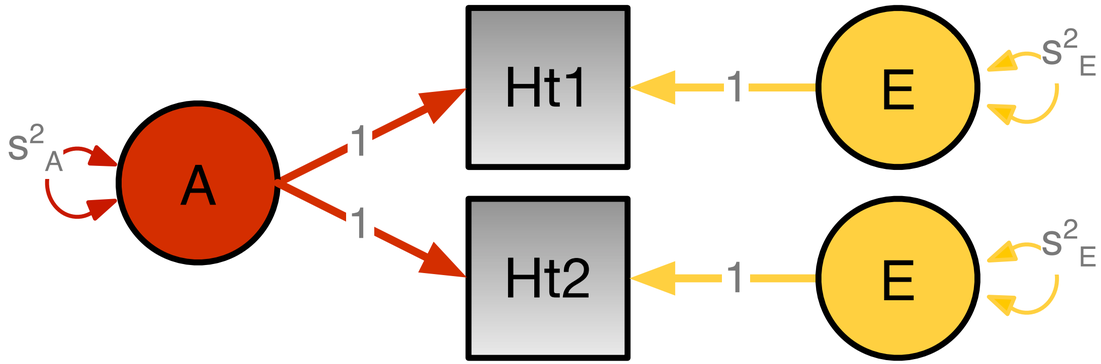
Polderman et al 2015 Nat Genet 47, 702–709 (2015). https://doi.org/10.1038/ng.3285

Estimate A & E

with MZT design: MZ twins raised together

- Hypothesis of interest: estimate s^2_A & s^2_E
  - A represents genetic effects s^2_A
  - E represents unshared environmental effects s^2_E



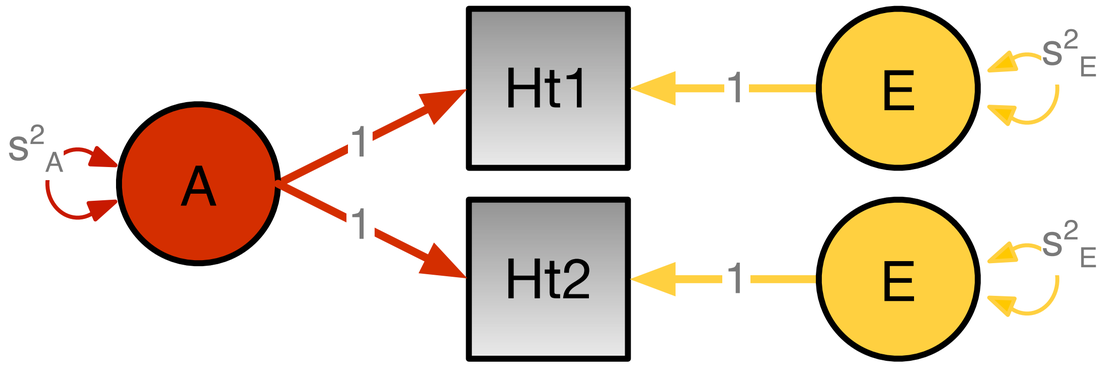


- Data: Height measured in 250 twin pairs (500 twins), sampling unit: twin pair
- **Key to GID**: MZ twins genetically identical, 100% genetic variance is shared, implies a covariance structure model; means to end of estimating s^2_A & s^2_E

Path Tracing II

with MZT design: MZ twins raised together

- 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



expected GSMS [Σ]		
	Height MZ1	Height MZ2
Ht MZ1	$s^2_A + s^2_E$	s^2_A
Ht MZ2	s^2_A	$s^2_A + s^2_E$

Contributions to Variance/Covariance

A & E

- Genes that contribute to height variance, necessarily contribute to MZ covariance, because MZ twins are genetically identical
  - Hypothesis: MZ1 variance $s^2_{Ht1} = \mathbf{s^2_A} + \mathbf{s^2_E}$ MZ2 variance $s^2_{Ht2} = \mathbf{s^2_A} + \mathbf{s^2_E}$

observed N=150 MZ pairs [S]		
	Height MZ1	Height MZ2
Ht MZ1	63.891	50.782
Ht MZ2	50.782	64.150

expected GSMS [Σ]		
	Height MZ1	Height MZ2
Ht MZ1	$\mathbf{s^2_A} + \mathbf{s^2_E}$	
Ht MZ2		$\mathbf{s^2_A} + \mathbf{s^2_E}$

- $\mathbf{s^2_A} = 50.782$ (estimate of **A** variance component)
- $\mathbf{s^2_E} = s^2_{Ph} - s^2_A = 63.89–50.78=13.11$ & $64.15–50.78=13.37 = (13.11+13.37)/2=13.24$ (estimate of **E**)

Contributions to Variance/Covariance

A & E

- Genes that contribute to height variance, necessarily contribute to MZ covariance, because MZ twins are genetically identical
  - Hypothesis: MZ1 variance $s^2_{Ht1} = \mathbf{s^2_A} + \mathbf{s^2_E}$ MZ2 variance $s^2_{Ht2} = \mathbf{s^2_A} + \mathbf{s^2_E}$
  - Hypothesis: MZ1-MZ2 covariance $s_{Ht1,Ht2} = \mathbf{s^2_A}$

observed N=150 MZ pairs [S]		
	Height MZ1	Height MZ2
Ht MZ1	63.891	50.782
Ht MZ2	50.782	64.150

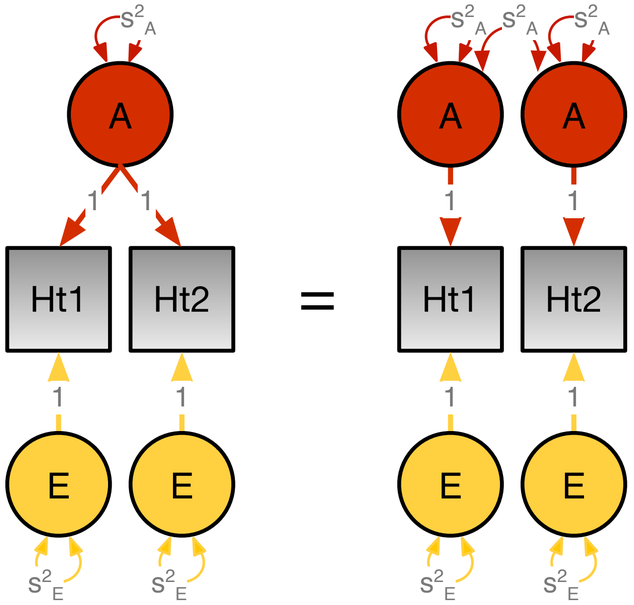
expected GSMS [Σ]		
	Height MZ1	Height MZ2
Ht MZ1	$\mathbf{s^2_A + s^2_E}$	$\mathbf{s^2_A}$
Ht MZ2	$\mathbf{s^2_A}$	$\mathbf{s^2_A + s^2_E}$

- $\mathbf{s^2_A} = 50.782$ (estimate of **A** variance component)
- $\mathbf{s^2_E} = s^2_{Ph} - s^2_A = 63.89–50.78=13.11$ & $64.15–50.78=13.37 = (13.11+13.37)/2=13.24$ (estimate of **E**)

Path Diagram - Equations - Cov Structures

GID= MZT

Graphically: Path Model



Regression Equations

- $Ht_1 = m + A_1 + E_1$
- $Ht_2 = m + A_2 + E_1$
- or $(A_1 = A_2)$

- $Ht_1 = m + A + E_1$
- $Ht_2 = m + A + E_2$

Covariance Structure

- variances &
- covariances

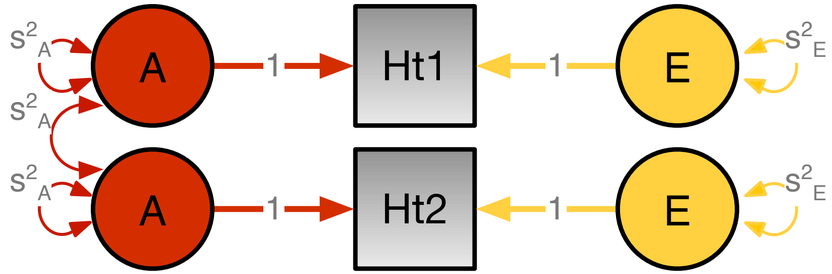
GSMS Model [Σ]		
	Ht1	Ht2
Ht1	$s^2_A + s^2_E$	s^2_A
Ht2	s^2_A	$s^2_A + s^2_E$

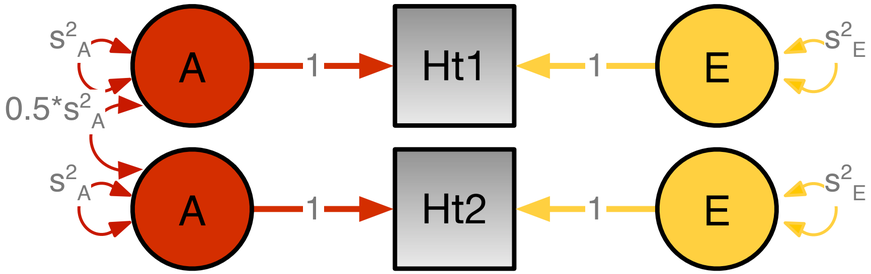
- MZT is a weak GID... what have we assumed concerning the environment? (see MZ1-MZ2 covariance!)

Classical Twin Design

MZ & DZ twins reared together

- MZ and DZ twins (raised/ growing up together in same household): AE model





Σ_{MZ}	Ht1	Ht2
Ht1	$s^2_A + s^2_E$	s^2_A
Ht2	s^2_A	$s^2_A + s^2_E$

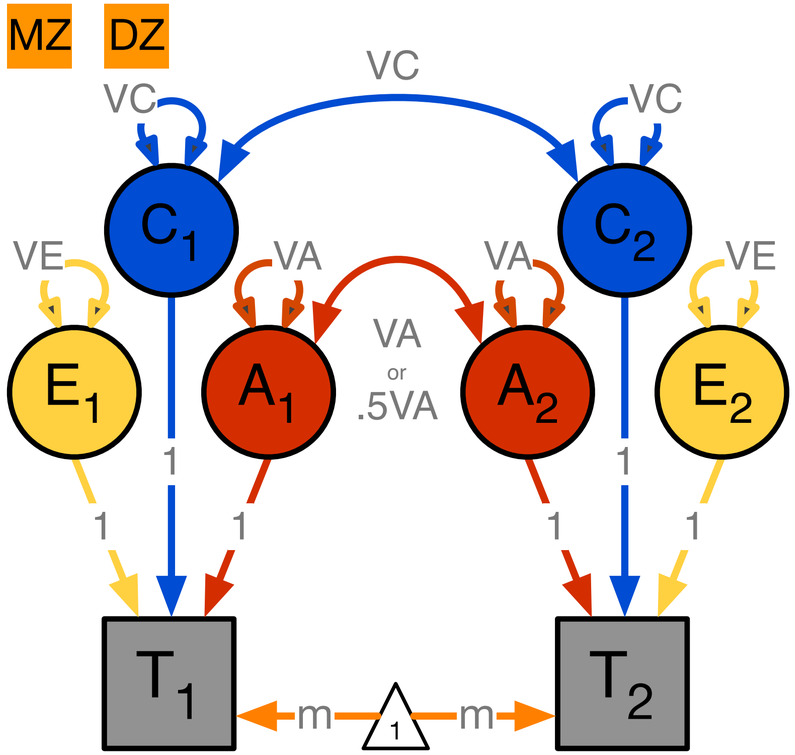
Σ_{DZ}	Ht1	Ht2
Ht1	$s^2_A + s^2_E$	$0.5*s^2_A$
Ht2	$0.5*s^2_A$	$s^2_A + s^2_E$

- Why add DZ twins? To extend the model (add variance component): ACE or ADE model

ACE Model

with CTD design

- **A** = additive genetic (rA=1 for MZ, 1/2 for DZ)
- **C**= common (shared) environmental
- **E** = unshared environmental
- (+ measurement error)



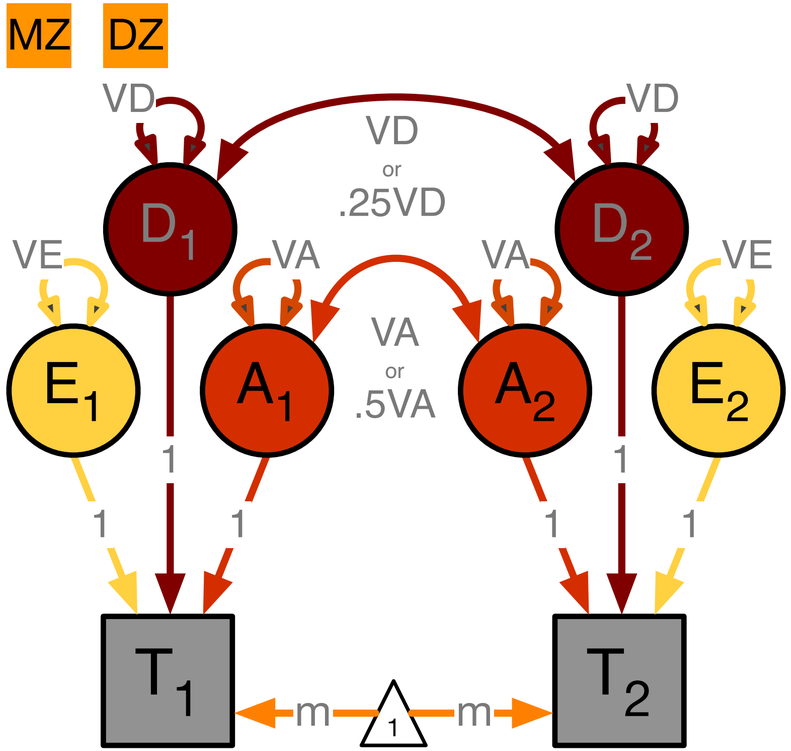
Σ_{MZ}	Ht1	Ht2
Ht1	$s^2_A + s^2_C + s^2_E$	$s^2_A + s^2_C$
Ht2	$s^2_A + s^2_C$	$s^2_A + s^2_C + s^2_E$

Σ_{DZ}	Ht1	Ht2
Ht1	$s^2_A + s^2_C + s^2_E$	$\frac{1}{2}*s^2_A + s^2_C$
Ht2	$\frac{1}{2}*s^2_A + s^2_C$	$s^2_A + s^2_C + s^2_E$

ADE Model

with CTD design

- **A** = additive genetic (rA=1 for MZ, 1/2 for DZ)
- **D**= additive genetic (rD=1 for MZ, 1/4 for DZ)
- **E** = unshared environmental
- (+ measurement error)



Σ_{MZ}	Ht1	Ht2
Ht1	$s^2_A + s^2_D + s^2_E$	$s^2_A + s^2_D$
Ht2	$s^2_A + s^2_D$	$s^2_A + s^2_D + s^2_E$

Σ_{DZ}	Ht1	Ht2
Ht1	$s^2_A + s^2_D + s^2_E$	$\frac{1}{2} * s^2_A + \frac{1}{4} s^2_D$
Ht2	$\frac{1}{2} * s^2_A + \frac{1}{4} s^2_D$	$s^2_A + s^2_D + s^2_E$

Univariate/ MonoPhenotype Twin Modeling in OpenMx

Hermine HM Maes with credit to Nick Martin, Elizabeth Prom-Wormley, Lindon Eaves, Tim Bates, Fruhling Rijsdijk, Brad Verhulst, Michael Neale & many others

Univariate Analysis - Classical Twin Design

roadmap for day 1

- Session 1:
  - Use data to test basic assumptions (equal means & variances for members of twin pairs): **Saturated Model**
- Session 2:
  - Estimate contributions of genetic/environmental effects to phenotypic variance: ACE or ADE Models
  - Test ACE / ADE submodels to identify significant genetic and environmental contributions: AE / (CE) / E Only Model
- Session 3:
  - Discuss Classical Twin Design [CTD] and other assumptions

Classical Twin Study Assumptions

focus on basic data assumptions

- 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

Classical Twin Study Assumptions

focus on basic data assumptions

- 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**
  - **Equal means/variances in MZ and DZ twins**
  - Further assumptions needed for male twins and opposite sex twin pairs

Practical Example

dataset: NH&MRC Twin Register 1981 Questionnaire

- **BMI** (body mass index): weight/height squared
  - kg/m^2, transformed: 7*log(BMI), simulated based on real data
- Young Female Cohort: 18-30 years; Sample Size:
  - MZf: 534 pairs (zyg=1; zygosity='MZFF' & cohort='younger')
  - DZf: 328 pairs (zyg=3; zygosity='DZFF' & cohort='younger')

```
> head(twinData)
  fam age zyg part wt1 wt2     ht1     ht2   htwt1   htwt2     bmi1     bmi2
1   1  21   1    2  58  57 1.7000 1.7000 20.0692 19.7232 20.9943 20.8726
2   2  24   1    2  54  53 1.6299 1.6299 20.3244 19.9481 21.0828 20.9519
3   3  21   1    2  55  50 1.6499 1.6799 20.2020 17.7154 21.0405 20.1210
4   4  21   1    2  66  76 1.5698 1.6499 26.7759 27.9155 23.0125 23.3043
5   5  19   1    2  50  48 1.6099 1.6299 19.2894 18.0662 20.7169 20.2583....
```

Observed Means, Variances, Covariances

BMI in young OZ female MZ & DZ twins

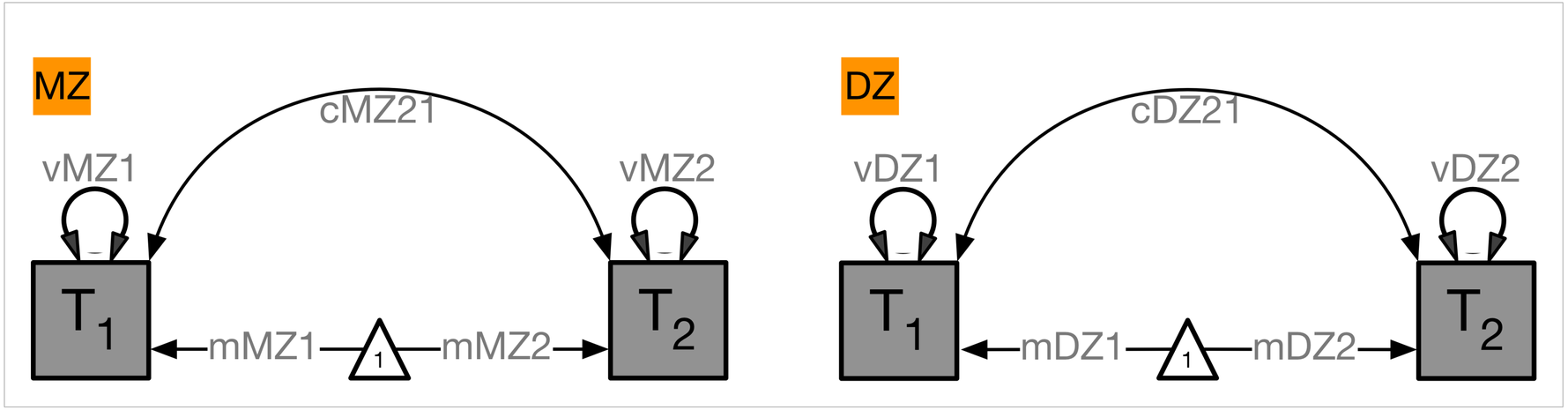
	MZ twins			DZ twins		
mean		**T1**	**T2**		**T1**	**T2**
	MZ	21.34	21.35	**DZ**	21.45	21.46
cov		**T1**	**T2**		**T1**	**T2**
	T1	0.73	0.59	**T1**	0.77	0.24
	T2	0.59	0.79	**T2**	0.24	0.82

- Lindon from his Hollywood Cemetery grave: ‘look at the bloody data’
- How can we actually be sure that the means and variances are truly the same?

Saturated Model

as many free parameters as statistics

- OpenMx Script: **oneSATc.R**
  - Saturated model estimating means & (co)variances for continuous data in MZ & DZ twins



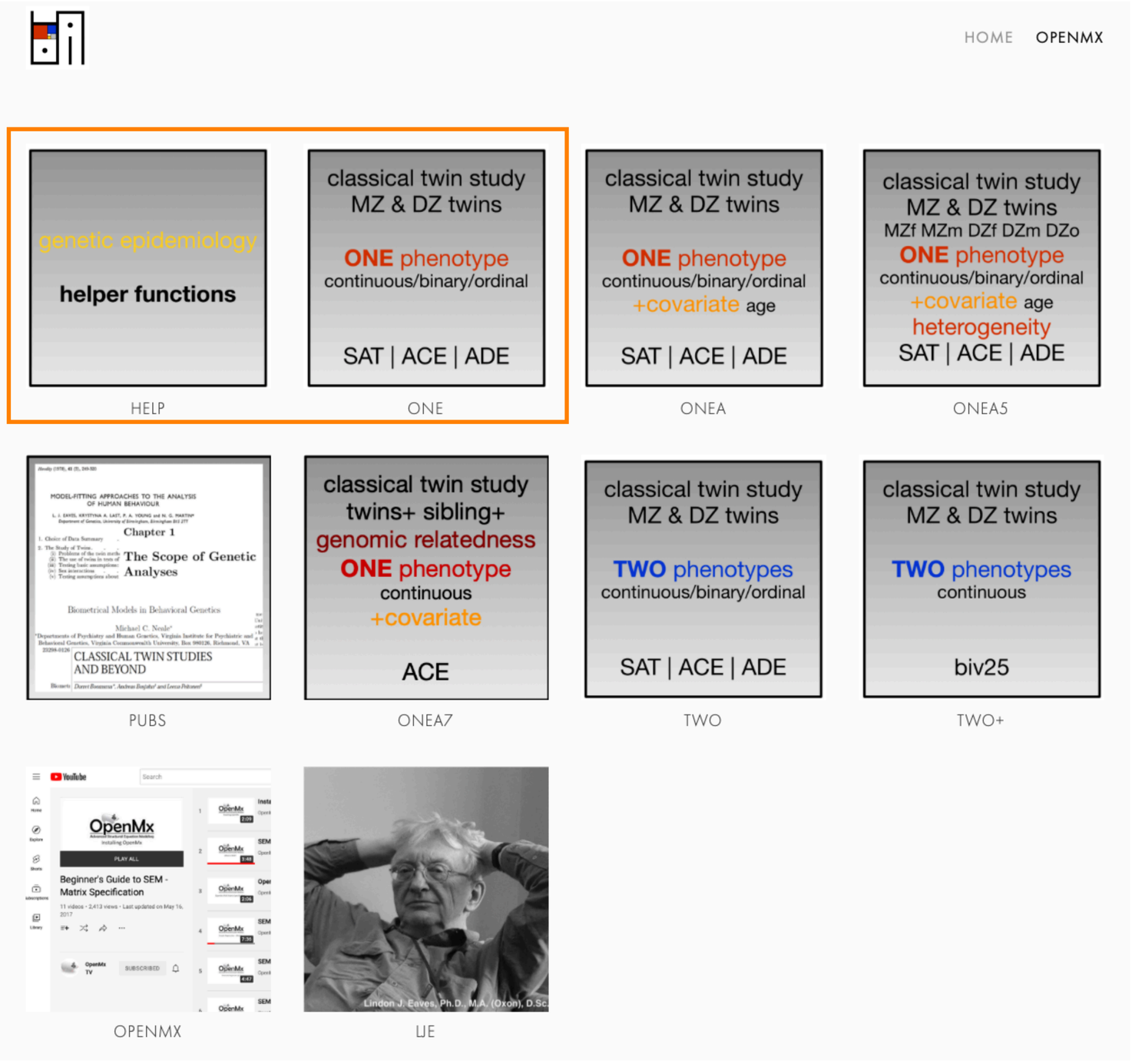
Intuition behind Maximum Likelihood (ML)

likelihood ratio tests

- Likelihood (L): probability that observation (data point) is predicted by model
- Maximum Likelihood Estimates (MLE): most likely values of population parameter values (e.g, μ, σ, β) given observed sample values
- Likelihood Ratio (LR): test statistic comparing Log-Likelihoods under 2 separate models: Mu: Unconstrained & Mc: Constrained (fewer parameters)
- LR statistic equals: $LR(Mc \mid Mu) = 2\ln(L(Mu) - 2\ln(L(Mc)$
- Likelihood Ratio Test (LRT) is asymptotically distributed as $\chi^2_{df=1}$ with degrees of freedom (df) equal to number of constraints (or parameters)
- If associated p-value < alpha (e.g. .05), we reject Null hypothesis

hermine-maes.squarespace.com

mifunctions.R

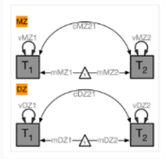
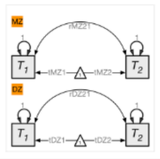
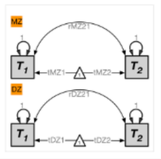
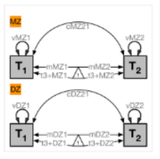
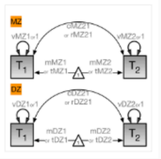
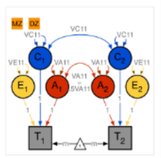
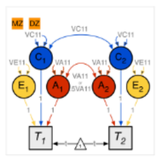
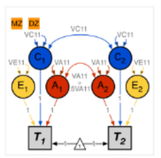
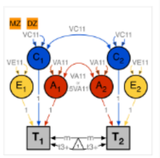
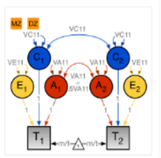
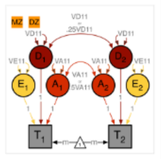
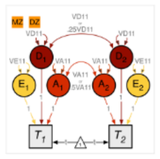
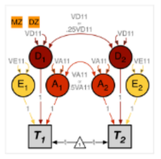
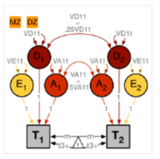
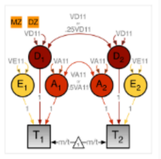


1 Phenotype
continuous

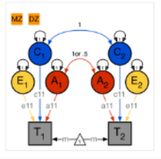
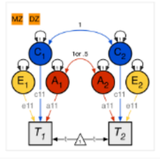
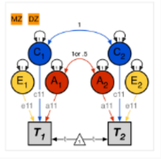
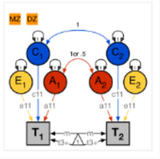
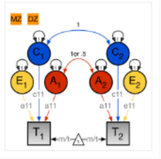
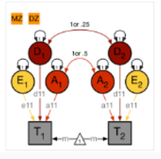
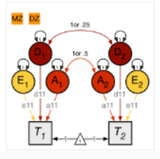
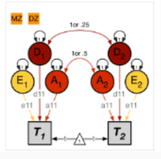
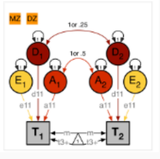
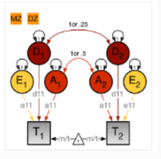
classical twin study
MZ & DZ twins

ONE phenotype
continuous/binary/ordinal

SAT | ACE | ADE

.	One Phenotype **CONTINUOUS** c	One Phenotype **BINARY** b	One Phenotype **ORDINAL** o	One Phenotype **ORDINAL** m *	One Phenotype **UMX** c/b/m
SAT estimating means/thresholds, variances & covariances * > 2 categories					
	oneSATc.R	**oneSATb.R**	**oneSATo.R**	**oneSATm.R**	**oneSATu.R**
ACE estimating variance components					
	oneACEvc.R	**oneACEvb.R**	**oneACEvo.R**	**oneACEvm.R**	**oneACEvu.R**
ADE estimating variance components					
	oneADEvc.R	**oneADEvb.R**	**oneADEvo.R**	**oneADEvm.R**	**oneADEvu.R**

NOTE! Models below estimating path coefficients may provide biased estimates of the parameters. Use of these scripts is discouraged.

ACE estimating path coefficients					
	oneACEc.R	**oneACEb.R**	**oneACEo.R**	**oneACEm.R**	**oneACEu.R**
ADE estimating path coefficients					
	oneADEc.R	**oneADEb.R**	**oneADEo.R**	**oneADEm.R**	**oneADEu.R**

Layout of my OpenMx Scripts

outline & naming conventions


```
Comments
Load Libraries & Options
Create Output
PREPARE DATA
Load Data
Select Variables for Analysis
Select Data for Analysis
Generate Descriptive Statistics
Set Starting Values
PREPARE MODEL
Create Algebra for expected Mean Matrices
Create Algebra for expected Variance/Covariance Matrices
Create Data Objects for Multiple Groups
Create Expectation Objects for Multiple Groups
Create Model Objects for Multiple Groups
Create Confidence Interval Objects
Build Saturated Model with Confidence Intervals
RUN MODEL
Run Saturated Model
Print Goodness-of-fit Statistics & Parameter Estimates
```

```
name of variable(s)          vars    <- 'bmi'
number of variables          nv      <- 1
number of twin variables     ntv     <- nv*2
variables per twin pair      selVars <- c('bmi1','bmi2')
definition variables         covVars
number of factors            nf      <- 2
number of thresholds         nth     <- 3

starting values              sv
lower/upper bound            lb / ub
labels                       lab

built model                  modelNAME
fitted model                 fitNAME
summary of fitted model      sumNAME
```

Saturated Model

oneSATc.R 1

load libraries, create output

oneSATc_isg2024.R

```
-------------------------------------------------------------------------------
Program: oneSATc.R
 Author: Hermine Maes
   Date: 10 22 2018
#
Twin Univariate Saturated model to estimate means and (co)variances across multiple groups
Matrix style model - Raw data - Continuous data
-------|---------|---------|---------|---------|---------|---------|---------|

Load Libraries & Options
rm(list=ls())
library(OpenMx)
library(psych); library(polycor)
source("miFunctions.R")

Create Output
filename    <- "oneSATc"
sink(paste0(filename,".Ro"), append=FALSE, split=TRUE)
```

load OpenMx

my functions which you can edit as you like

creates output file with extension .Ro

Preparing Data

oneSATc.R 2

load data, select variables/data, set values

```
Load Data
data(twinData)
dim(twinData); describe(twinData[,1:12], skew=F)

Select Variables for Analysis
vars      <- 'bmi'                        list of variables names
nv        <- 1                            number of variables
ntv       <- nv*2                         number of total variables
selVars   <- paste(vars,c(rep(1,nv),rep(2,nv)),sep="")

Select Data for Analysis
mzData    <- subset(twinData, zyg==1, selVars)
dzData    <- subset(twinData, zyg==3, selVars)

Generate Descriptive Statistics ..R colmeans/cov

Set Starting Values
svMe      <- 20                           start value for means
svVa      <- .8                           start value for variance
lbVa      <- .0001                        lower bound for variance
```

load ‘twinData’ or read in your own

analyzing c(‘bmi1’,’bmi2’)

zygosity=‘MZFF’ & cohort=‘younger’ → get right codes for zygosity

SAT Deconstructed

specify means & covariance Matrices


```
meanMZ     <- mxMatrix( type="Full", nrow=1, ncol=ntv,
 free=TRUE, values=svMe, labels=c("mMZ1","mMZ2"),name="meanMZ" )
meanDZ     <- mxMatrix( type="Full", nrow=1, ncol=ntv,
 free=TRUE, values=svMe, labels=c("mDZ1","mDZ2"),name="meanDZ" )
```

meanMZ *1x2*

mMZ1	mMZ2

meanDZ *1x2*

mDZ1	mDZ2

```
covMZ      <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv,
 free=TRUE, values=valDiag(svVa,ntv), lbound=valDiag(lbVa,ntv),
 labels=c("vMZ1","cMZ21","vMZ2"), name="covMZ" )
covDZ      <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv,
 free=TRUE, values=valDiag(svVa,ntv), lbound=valDiag(lbVa,ntv),
 labels=c("vDZ1","cDZ21","vDZ2"), name="covDZ" )
```

covMZ *2x2*

vMZ1	cMZ21
cMZ21	vMZ2

covDZ *2x2*

vDZ1	cDZ21
cDZ21	vDZ2

Preparing Model

oneSATc.R 3

expected means, (co)variances, data & expectation objects

```
Create Algebra for expected Mean Matrices
meanMZ     <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=svMe, labels=c("mMZ1","mMZ2"), name="meanMZ" )
meanDZ     <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=svMe, labels=c("mDZ1","mDZ2"), name="meanDZ" )

Create Algebra for expected Variance/Covariance Matrices
covMZ      <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=TRUE, values=valDiag(svVa,ntv), lbound=valDiag(lbVa,ntv),
labels=c("vMZ1","cMZ21","vMZ2"), name="covMZ" )
covDZ      <- mxMatrix( type="Symm", nrow=ntv, ncol=ntv, free=TRUE, values=valDiag(svVa,ntv), lbound=valDiag(lbVa,ntv),
labels=c("vDZ1","cDZ21","vDZ2"), name="covDZ" )

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()
```

full 1x2 matrix for means

symmetric 2x2 matrix for covariances

function in miFunctions.R
to put a value on all diagonal elements

fitting to raw data

link model to data

using FIML: full information maximum likelihood

Run Model

oneSATc.R 4

model objects, CIs, build & run model, print output

model object contains all matrices etc.

```
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") )
```

evaluating 2 groups simultaneously

```
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 )
```

built model

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

fitted model

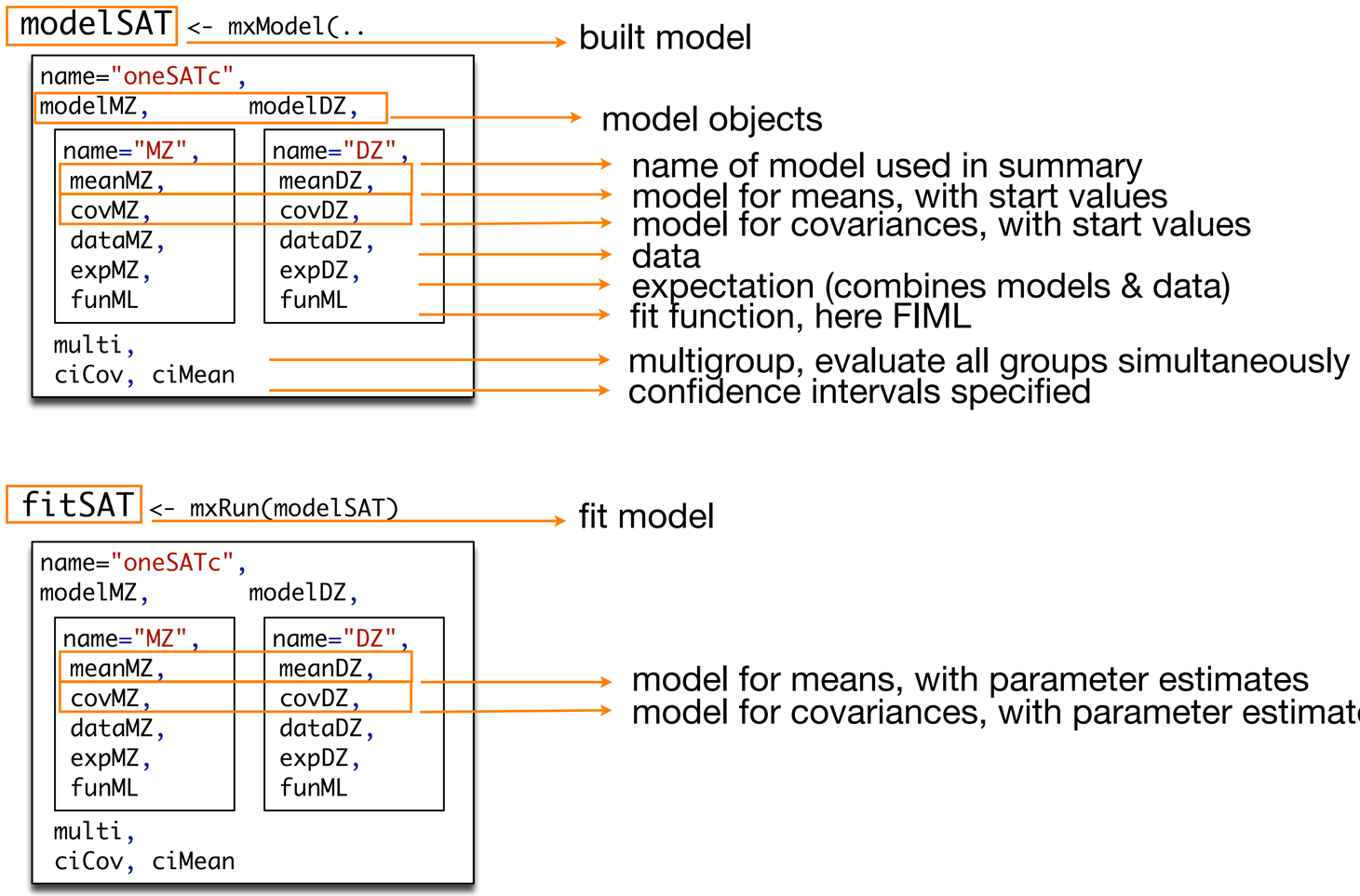
standard summary function in OpenMx

```
Print Goodness-of-fit Statistics & Parameter Estimates
fitGofs(fitSAT)
fitEsts(fitSAT)
mxGetExpected( fitSAT, c("means","covariance") )
```

my short summary functions in miFunctions.R

Model Building - Model Fitting

modelSAT vs fitSAT



Print Goodness-of-Fit Statistics

summary(fitSAT) vs fitGofs(fitSAT)

```
> summary(fitSAT)
Model Statistics:
               |  Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
       Model:              10                  1767             4055.9346
   Saturated:              NA                    NA                    NA
Independence:              NA                    NA                    NA
Number of observations/statistics: 920/1777

Information Criteria:
      |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
AIC:       521.93461             4075.9346                       NA
BIC:    -8002.73367             4124.1783                 4092.4195
CFI: NA
TLI: 1   (also known as NNFI)
RMSEA:  0  [95% CI (NA, NA)]
Prob(RMSEA <= 0.05): NA

> fitGofs(fitSAT)
Mx:oneSATc  os=1777  ns=920  ep=10  co=0  df=1767  ll=4055.9346  cpu=0.0952  opt=NPSOL  ver=2.17.2  stc=0
```

Print Estimates

summary(fitSAT)$parameters vs fitEsts(fitSAT)

```
> summary(fitSAT)$parameters
free parameters:
   name    matrix  row  col    Estimate   Std.Error A lbound ubound
1   mMZ1 MZ.meanMZ    1 bmi1 21.34437690 0.036061832
2   mMZ2 MZ.meanMZ    1 bmi2 21.34901242 0.037650856
3   vMZ1  MZ.covMZ bmi1 bmi1  0.72766891 0.043658984    1e-04
4  cMZ21  MZ.covMZ bmi1 bmi2  0.59163768 0.040794161        0
5   vMZ2  MZ.covMZ bmi2 bmi2  0.79319915 0.047647906    1e-04
6   mDZ1 DZ.meanDZ    1 bmi1 21.44752035 0.047571928
7   mDZ2 DZ.meanDZ    1 bmi2 21.45784215 0.049233334
8   vDZ1  DZ.covDZ bmi1 bmi1  0.76919130 0.059007266    1e-04
9  cDZ21  DZ.covDZ bmi1 bmi2  0.24004049 0.045201541        0
10  vDZ2  DZ.covDZ bmi2 bmi2  0.82163163 0.063154677    1e-04

> fitEsts(fitSAT)
    mMZ1     mMZ2     vMZ1    cMZ21     vMZ2     mDZ1     mDZ2     vDZ1    cDZ21     vDZ2
21.3444 21.3490  0.7277  0.5916  0.7932 21.4475 21.4578  0.7692  0.2400  0.8216
```

Estimated Values & Goodness of Fit Stats

saturated Model

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

10 parameters estimated:

`mMZ1, mMZ2, vMZ1, vMZ2, cMZ21`

`mDZ1, mDZ2, vDZ1, vDZ2, cDZ21`

	os	**ep**	**-2ll**	**df**	**AIC**
saturated	1777	10	4055.93	1767	521.93

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

Fitting Nested Models

oneSATc.R 5

equal means by twin order

changing parameters

```
Constrain expected Means to be equal across twin order
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' )
fitEMO    <- mxRun( modelEMO, intervals=F )
fitGofs(fitEMO); fitEsts(fitEMO)

Print Comparative Fit Statistics
mxCompare( fitSAT, fitEMO )
```

existing parameters

new parameters

generate likelihood ratio test

Estimated Values & Goodness of Fit Stats

means equated by twin order

	MZ twins			DZ twins		
mean		**T1**	**T2**		**T1**	**T2**
	MZ	21.35	21.35	**DZ**	21.45	21.45
cov		**T1**	**T2**		**T1**	**T2**
	T1	0.73		**T1**	0.77	
	T2	0.59	0.79	**T2**	0.24	0.82

8 parameters estimated:

`mMZ, vMZ1, vMZ2, cMZ21`

`mDZ, vDZ1, vDZ2, cDZ21`

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

	os	ep	-2ll	df	AIC	Δ-2ll	Δdf	p
saturated	1777	10	4055.93	1767	521.93			
mT1=mT2	1777	8	4056.00	1769	518.00	0.07	2	0.97

Practical II
Mean/Variance Estimation

- system("mkdir $HOME/hermine")
- system("cp -R /faculty/hmaes/2024/isg1.4/* $HOME/hermine")
- setwd("~hermine")

- `oneREGc_isg2024.R`

- OR `oneSATc.R` & `miFunctions.R` from https://hermine-maes.squarespace.com

ISG Workshop 2024

Fitting Nested Models

oneSATc.R 6

equal means/variances by twin order/zygosity

```
Constrain expected Means and Variances to be equal across twin order
modelEMVO <- mxModel(fitEMO, name="oneEMVOc" )
modelEMVO <- omxSetParameters( modelEMVO, label=c("vMZ1","vMZ2"), free=TRUE, values=svVa, newlabels='vMZ' )
modelEMVO <- omxSetParameters( modelEMVO, label=c("vDZ1","vDZ2"), free=TRUE, values=svVa, newlabels='vDZ' )
fitEMVO   <- mxRun( modelEMVO, intervals=F )
fitGofs(fitEMVO); fitEsts(fitEMVO)

Constrain expected Means and Variances to be equal across twin order and zygosity
modelEMVZ <- mxModel(fitEMVO, name="oneEMVZc" )
modelEMVZ <- omxSetParameters( modelEMVZ, label=c("mMZ","mDZ"), free=TRUE, values=svMe, newlabels='mZ' )
modelEMVZ <- omxSetParameters( modelEMVZ, label=c("vMZ","vDZ"), free=TRUE, values=svVa, newlabels='vZ' )
fitEMVZ   <- mxRun( modelEMVZ, intervals=F )
fitGofs(fitEMVZ); fitEsts(fitEMVZ)

Print Comparative Fit Statistics
mxCompare( fitSAT, subs <- list(fitEMO, fitEMVO, fitEMVZ) )

sink()
save.image(paste0(filename,".Ri"))
```

close .Ro file & save image as file with .Ri extension

Estimated Values & Goodness of Fit Stats

means & variances equated by twin order & zygosity

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

4 parameters estimated:

`mZ, vZ, cMZ21, cDZ21`

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

	os	ep	-2ll	df	AIC	Δ-2ll	Δdf	p
saturated	1777	10	4055.93	1767	521.93			
mT1=mT2	1777	8	4056.00	1769	518.00	0.07	2	0.97
+ varT1=varT2	1777	6	4058.94	1771	516.94	3.01	4	0.56
+ Zyg MZ=DZ	1777	4	4063.45	1773	517.45	7.52	6	0.28

Conclusions

basic data assumptions

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

Thank you

OpenMx Development Team (most active ones)

Steve Boker

Mike Neale

Tim Bates

umx functions to run

saturated & ACE models

Mike Hunter

Tim Brick

Rob Kirkpatrick

Joshua Pritikin

1987-2024

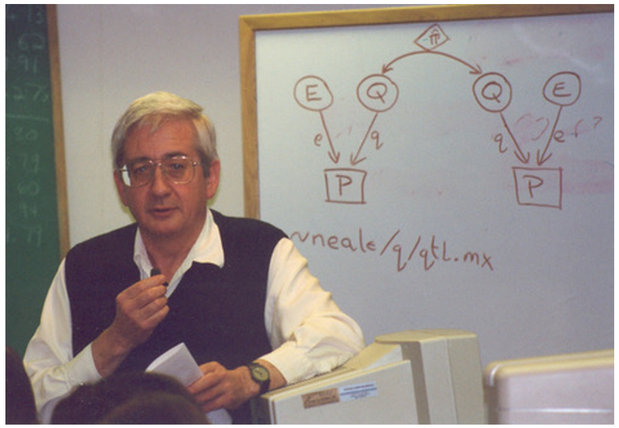
38th workshop

Thank you

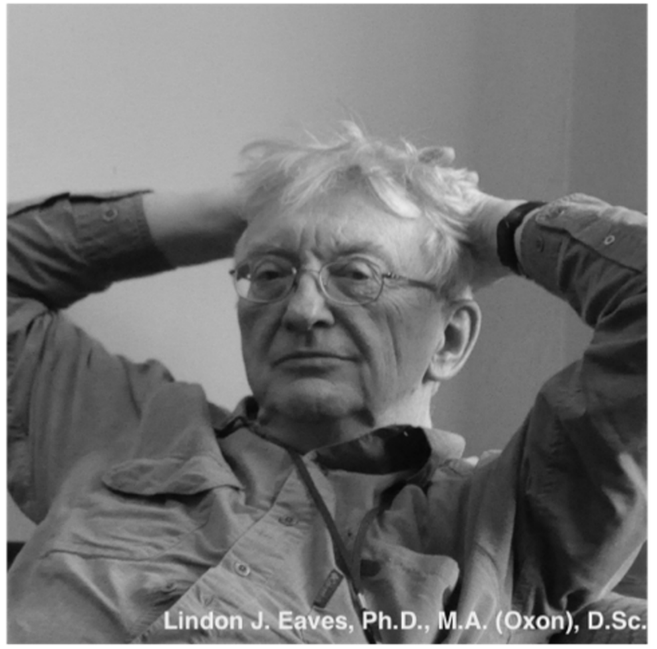
Thank you

Thank you

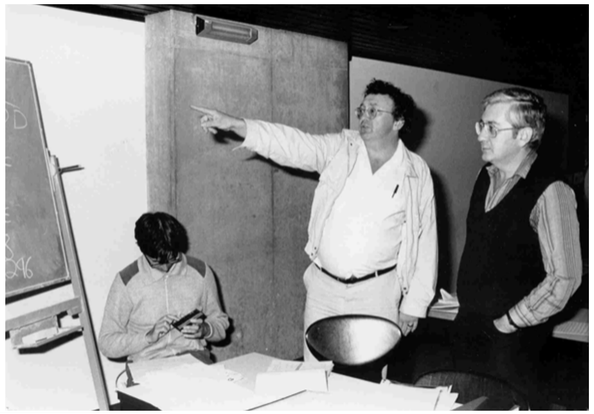
1989 Leuven workshop



David Fulker



Lindon Eaves



2005 Boulder 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† [workshop founders]

Advantages of GCSM in GIDs

GCSM also referred to as Structural Equation Modeling SEM

- Optimal estimates of parameters + information about precision of estimates via 95% confidence intervals (CIs)
- Overall goodness of fit testing: does model fit observed covariance matrices?
- Statistical testing of individual parameters (ACE vs AE; ACE vs CE; ADE vs AE)
- Generalizes from 1 phenotype to P phenotypes (multivariate phenotype / repeated measures)
- Accommodates missing data
- Can handle binary / dichotomous phenotypic data
- Can handle any (multivariate) Genetically Informative Design (e.g. twins + parents; twins + siblings; children of twin design; extended pedigree design)

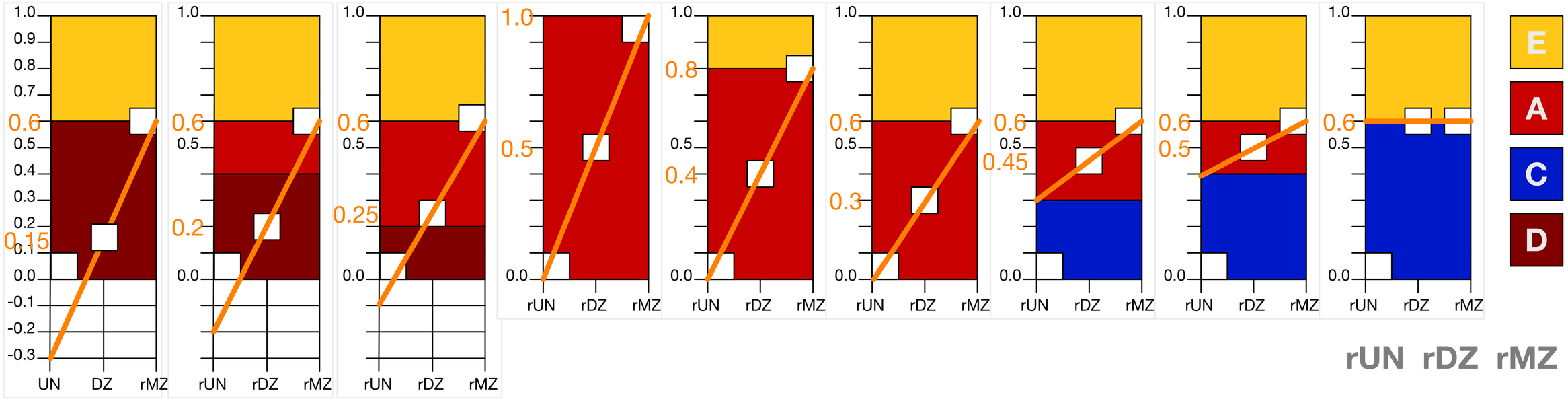
GCSM/SEM

We have

- Data (MZ and DZ twin phenotypic data)
- Data summary (linear relationship): two covariance matrices & 4 means
- Linear model for the data (path model), which implies covariance structure(s)
- Covariance structure(s): covariance matrices expressed in terms of unknown (to be estimated) and known parameters (GID!)
- Classical twin design: unknown parameters (s^2_A , s^2_C, s^2_E) and known parameters (1, ½, ¼)
- How to obtain estimates? Maximum likelihood estimation
- Parameters associated with hypothesized ACE model
  - **μ** phenotypic mean; s^2_A genetic; s^2_C shared env; s^2_E unshared env variance

From Twin Correlations

to Sources of Variance: A, C, D & E



rUN rDZ rMZ

rDZ < 1/2 rMZ

E + A & D

D= -2*C / A'= A+3C

rMZ=2*rDZ [rMZ > rDZ, 1–rMZ >0]

only A

E + only A

rDZ > 1/2rMZ

E + A & C

C= -1/2*D / A'= A +3/2D

rMZ = rDZ

E + only C

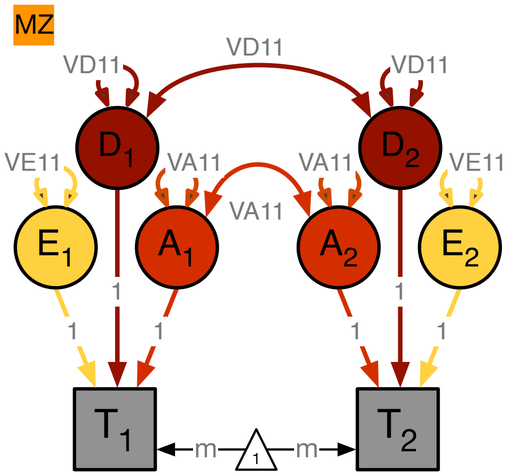
rMZ: monozygotic twin correlation; rDZ: dizygotic twin correlation; rUN: unrelated correlation

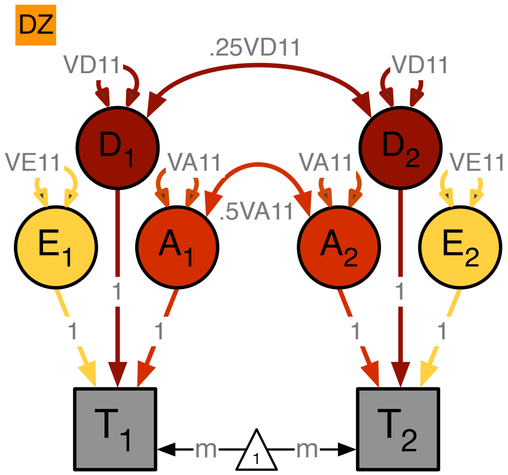
A: additive genetic variance; E: unique environmental variance: C: common environmental variance; D: dominance genetic variance

Genetic Model(s)

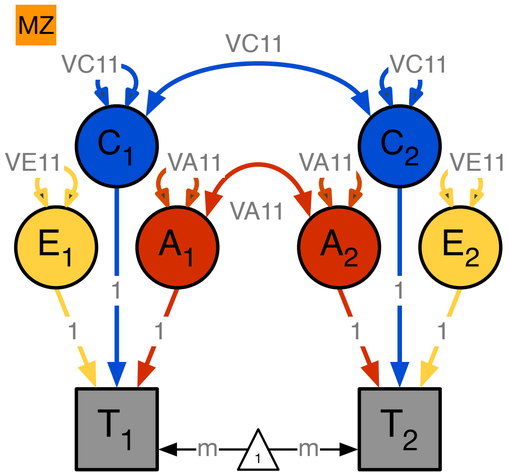
ADE & ACE Models

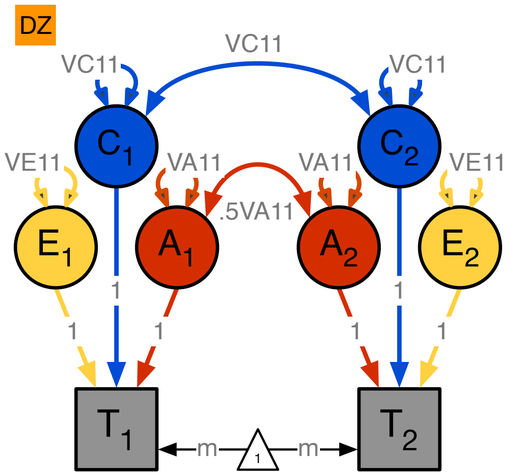
- oneADEvc.R





oneACEvc.R





Estimating ACE Model

- 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
  - cMZ – cDZ = (VA+VC) – (.5VA+VC) = .5VA, thus VA = 2(cMZ – cDZ)
  - VC = –cMZ +2cDZ, or V – VA – VE = VC
- What if negative VC?
  - VD' = –2VC
  - VA' = cMZ +2VC = VA+3VC

Estimating ADE Model

- 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, thus VA = –(cMZ – 4cDZ)
  - VD = 2cMZ – 4cDZ, or V – VA – VE = VD
- What if negative VD?
  - VC' = –1/2VD
  - VA' = cMZ +1/2VD = VA+3/2VD