Univariate modeling

Sarah Medland

Starting at the beginning…

- Data preparation
  - The algebra style used in Mx expects 1 line per case/family
  - (Almost) limitless number of families and variables
  - Missing data
    - Default missing code is **NA**
    - **No missing covariates/definition variables!**
  - Quick R - http://www.statmethods.net/

Selecting and sub-setting data

- Make separate data sets for the MZ and DZ

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

- Check data is numeric and behaves as expected

  cov (dzData, use="complete")

Common problem

- Problem: data contains a non numeric value

```
> cov(dzData,use="complete")
          bmi1      bmi2
bmi1 0.8908474 0.2872594
bmi2 0.2872594 0.8657751
Warning message:
In cov(dzData, use = "complete") : NAs introduced by coercion
> colMeans(mzData,na.rm=TRUE)
Error in colMeans(mzData, na.rm = TRUE) : 'x' must be numeric
> colMeans(dzData,na.rm=TRUE)
Error in colMeans(dzData, na.rm = TRUE) : 'x' must be numeric
```

20171	1	0.35	5	2	51	79	1.5999	1.7998	19.9219	24.3827	20.9427	22.3571
20188	1	0.37	5	2	53	65	1.5698	1.73	21.5019	21.7181	21.477	21.547
20204	1	0.53	5	1	58	64	1.6299	NA	21.83	NA	A	NA
20390	1	0.37	5	2	64	73	1.6499	1.8298	23.5078	21.7982	22.1013	21.5728
20398	1	0.52	5	2	60	77	1.6299	1.73	22.5827	25.7276	21.8203	22.7329

- Equivalent Mx Classic error - *Uh-oh... I'm having trouble reading a number in D or E format*

Important structural stuff

- openMx has a very fluid and flexible stucture
- Each code snippet is being saved as an object
- We tend to reuse the object names in our scripts
- This makes it very important to create a new project for each series of analyses
- Remember the project also contains the data so these files can become very large.

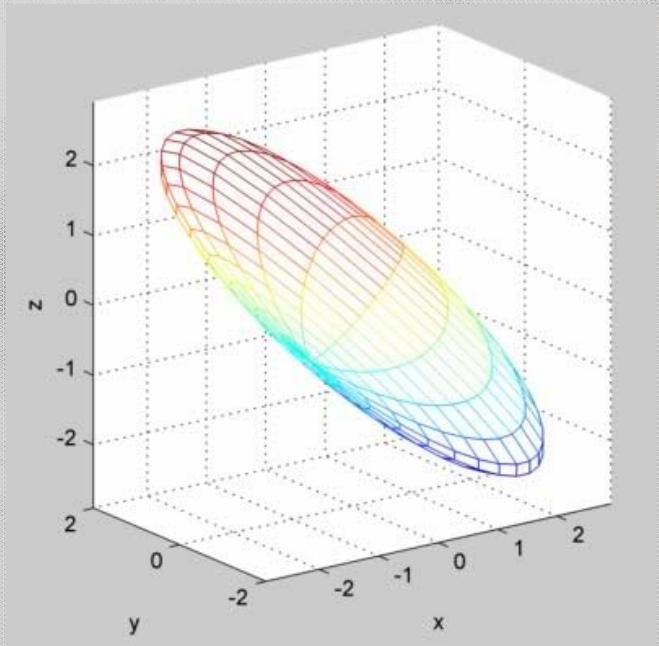
Matrices are the building blocks

```
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE,
values=.6, label="a11", name="a" ), #X
```

- Many types eg. type="Lower"
- Denoted by names eg. name="a“
- Size eg. nrow=nv, ncol=nv
- All estimated parameters must be placed in a matrix & Mx must be told what type of matrix it is

Choosing the model

- Thinking about parameter space…
- Imagine an ACE model
- Solution space bounded by CIs



Choosing the model

- ACE vs ADE
  - With twins alone can't joint estimate ACDE
  - Options
    - Add in an extra relationship
    - Fix one of these parameters and estimate the other 3
    - Accept this limitation
      - All models are wrong some are useful (George E. P. Box)
    - Reject the twin model, pretend genes have no influence and interpret biological inheritance as a social phenomenon
  - No 1 size fits all solution

Quantifying and Addressing Parameter Indeterminacy in the Classical Twin Design

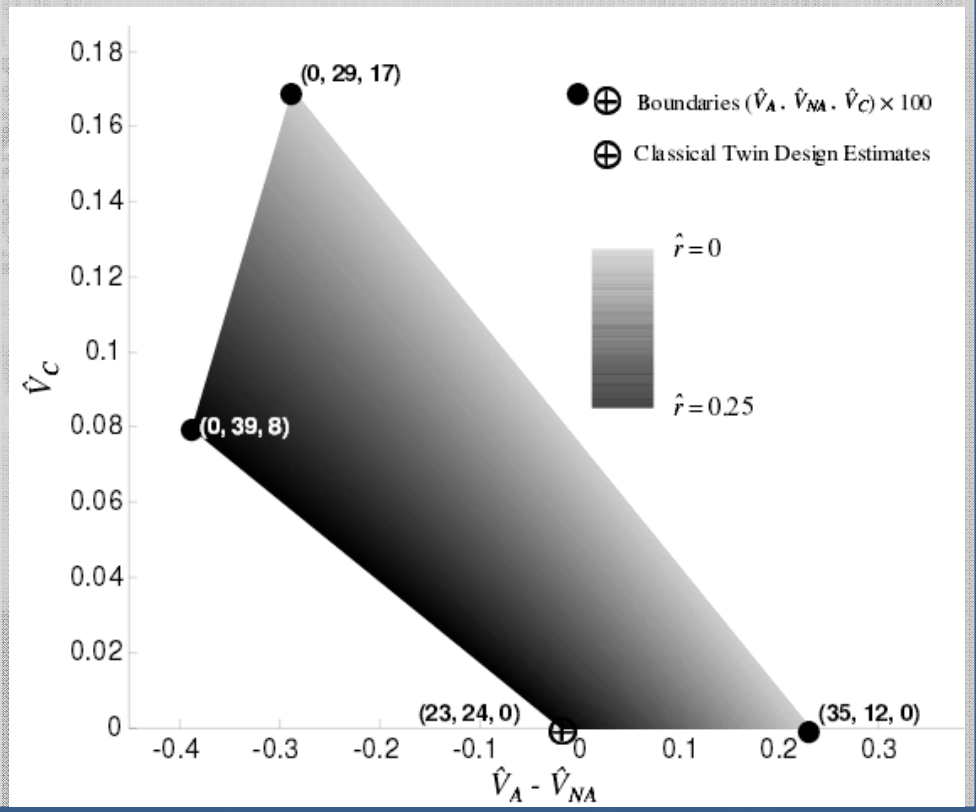
Matthew C. Keller[1] and William L. Coventry[2,3]

[1]Center for Society and Genetics, University of California, Los Angeles, United States of America
[2]Queensland Institute of Medical Research, Brisbane, Australia
[3]School of Psychology, University of New England, Armidale, Australia

Table 1

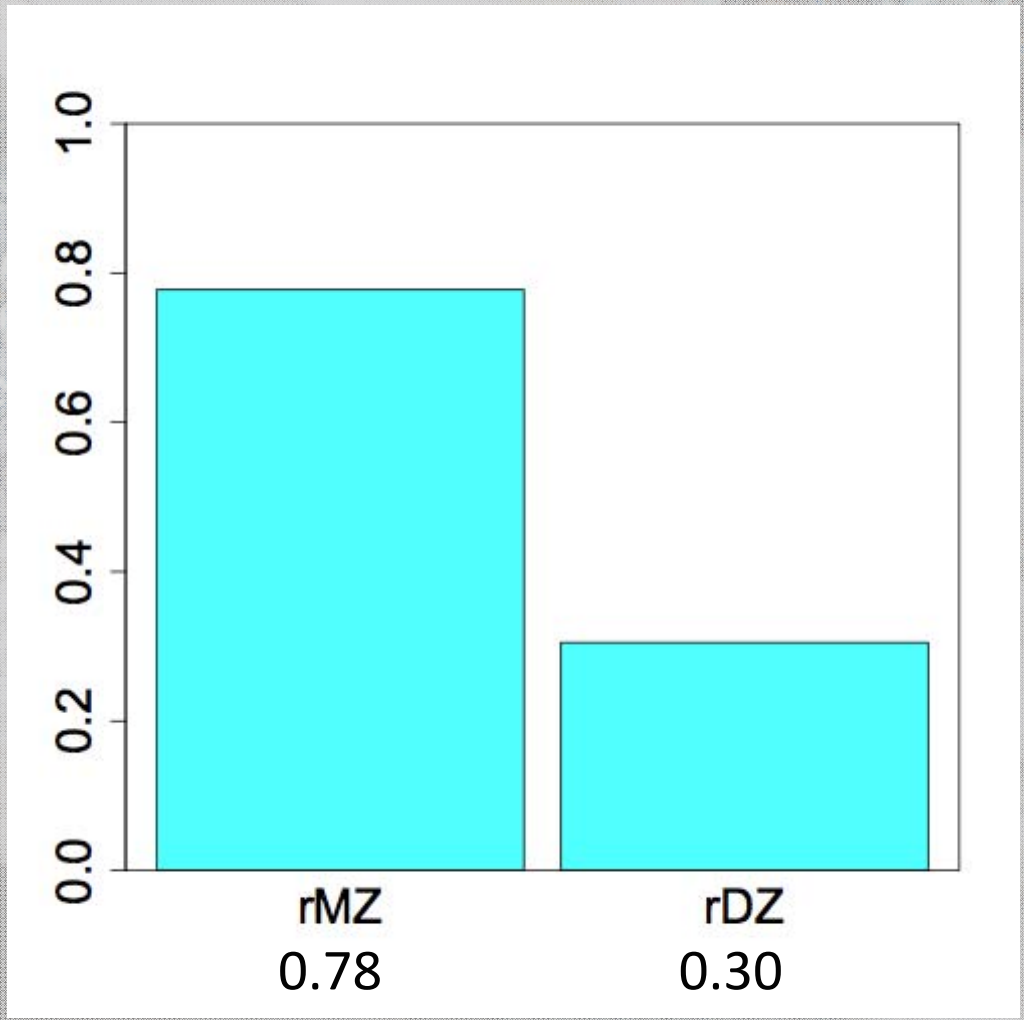
Methods of Obtaining $\hat{V}_A$, $\hat{V}_C$, and $\hat{V}_{NA}$ from $C\hat{V}_{NZ}$ and $C\hat{V}_{OZ}$ and the Boundaries of the Parameter Space Given Eight Different Pairs of Fixed Parameters Possible in Twin-Only Designs

Fixed Parameters	$\hat{V}_A$	$\hat{V}_{NA}$	$\hat{V}_C$
$C\hat{V}_{OZ}/C\hat{V}_{NZ} > 1/2$			
1. $\hat{r}=0$, $\hat{V}_A=0$	0 min	$C\hat{V}_{NZ}-C\hat{V}_{OZ}$ inter	$C\hat{V}_{OZ}$ max
2. $\hat{r}=0$, $\hat{V}_{NA}=0$	$2(C\hat{V}_{NZ}-C\hat{V}_{OZ})$ max	0 min	$2C\hat{V}_{OZ}-C\hat{V}_{NZ}$ min
3. $\hat{r}=.25$, $\hat{V}_A=0$	0 min	$4/3(C\hat{V}_{NZ}-C\hat{V}_{OZ})$ max	$4/3C\hat{V}_{OZ}-1/3C\hat{V}_{NZ}$ inter
4. $\hat{r}=.25$, $\hat{V}_{NA}=0$	$2(C\hat{V}_{NZ}-C\hat{V}_{OZ})$ max	0 min	$2C\hat{V}_{OZ}-C\hat{V}_{NZ}$ min
$C\hat{V}_{OZ}/C\hat{V}_{NZ} < 1/2$			
5. $\hat{r}=0$, $\hat{V}_A=0$	0 min	$C\hat{V}_{NZ}-C\hat{V}_{OZ}$ inter	$C\hat{V}_{OZ}$ max
6. $\hat{r}=0$, $\hat{V}_C=0$	$2C\hat{V}_{OZ}$ max	$C\hat{V}_{NZ}-2C\hat{V}_{OZ}$ min	0 min
7. $\hat{r}=.25$, $\hat{V}_A=0$	0 min	$4/3(C\hat{V}_{NZ}-C\hat{V}_{OZ})$ max	$4/3C\hat{V}_{OZ}-1/3C\hat{V}_{NZ}$ inter
8. $\hat{r}=.25$, $\hat{V}_C=0$	$4C\hat{V}_{OZ}-C\hat{V}_{NZ}$ inter	$2C\hat{V}_{OZ}-4C\hat{V}_{NZ}$ inter	0 min



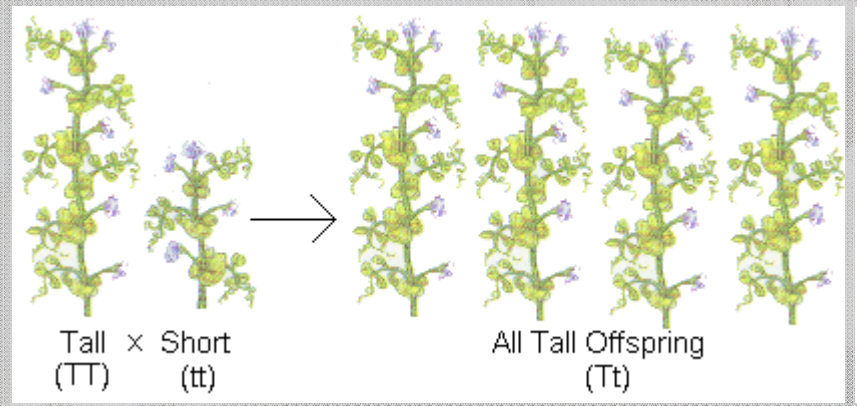
Yesterday we ran an ADE Model

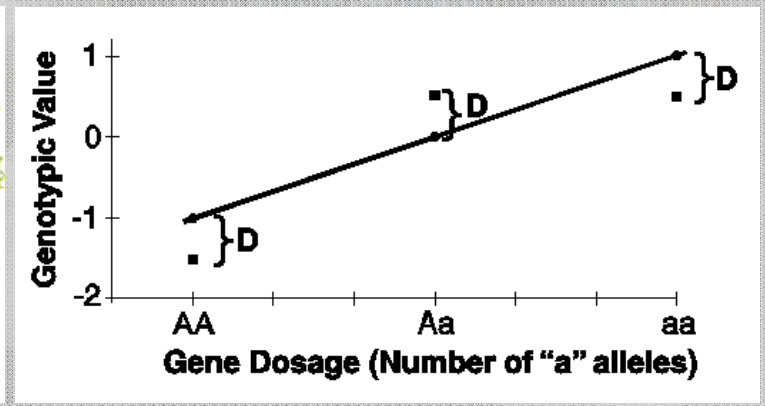
- Why?



What is D again?

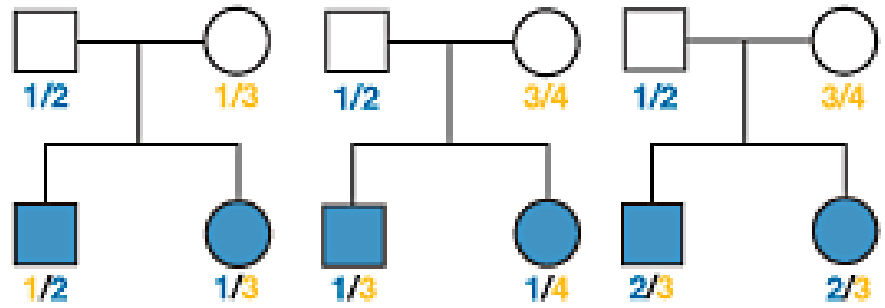
- Dominance refers to non-additive genetic effects resulting from interactions between alleles at the same locus or different loci (epistasis)





What is D again?

- DZ twins/full siblings share
  - ~50% of their segregating DNA &
  - for ~25% loci they share not only the genotype but also the parental origin of each allele



- DZ twins/full siblings share
  - ~50% of their segregating DNA &
  - for ~25% loci they share not only the genotype but also the parental origin of each allele

This is where the .5A comes from

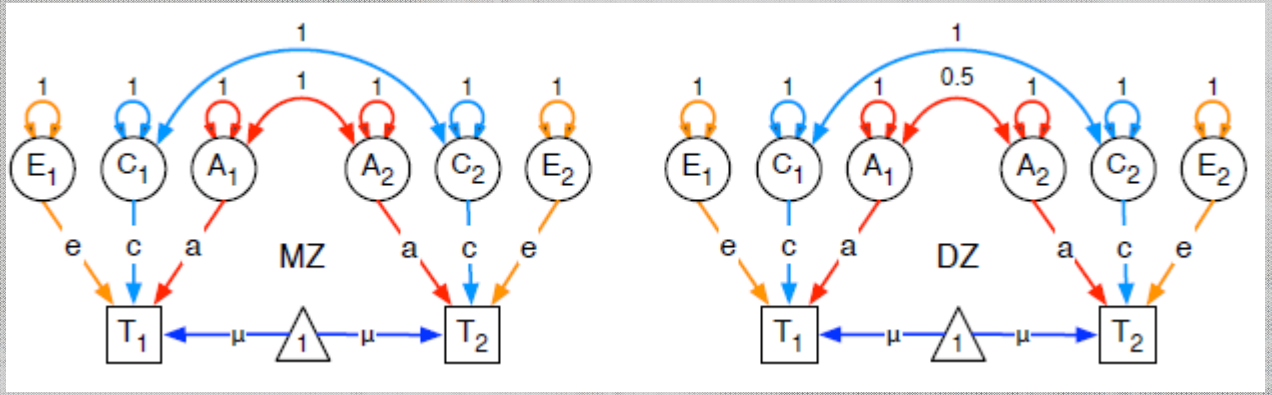
This is where the .25D comes from

Consider a mating between mother AB x father CD:

		Sib1			
		AC	**AD**	**BC**	**BD**
Sib 2	**AC**	2	1	1	0
	AD	1	2	0	1
	BC	1	0	2	1
	BD	0	1	1	2

IBD 0 : 1 : 2 = 25% : 50% : 25%

Today we will run an ACE model

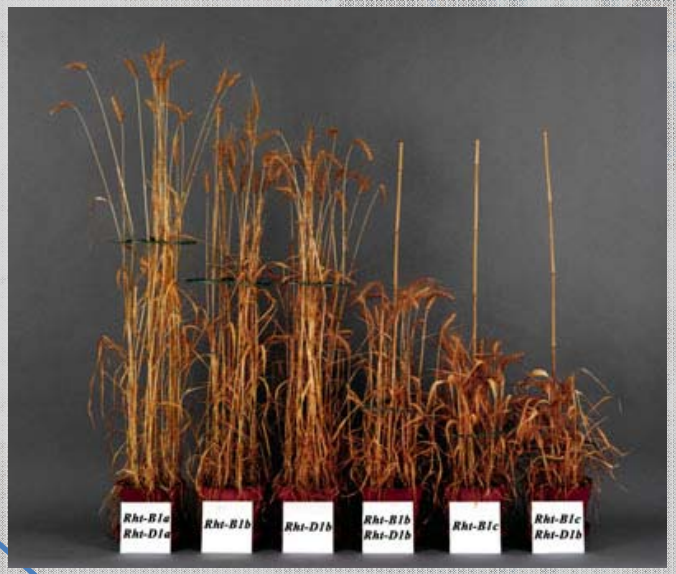


MZ		DZ	
$a^2+c^2+e^2$	a^2+c^2	$a^2+c^2+e^2$	$.5a^2+c^2$
a^2+c^2	$a^2+c^2+e^2$	$.5a^2+c^2$	$a^2+c^2+e^2$

Today we will run an ACE model

Additive genetic effects

- Why is the coefficient for DZ pairs .5?
- Average genetic sharing between siblings/DZ twins



Sib 2 \ Sib1	AC	AD	BC	BD
AC	2	1	1	0
AD	1	2	0	1
BC	1	0	2	1
BD	0	1	1	2

MZ

$a^2+c^2+e^2$ a^2+c^2

a^2+c^2 $a^2+c^2+e^2$

DZ

$a^2+c^2+e^2$ $.5a^2+c^2$

$.5a^2+c^2$ $a^2+c^2+e^2$

Today we will run an ACE model

Common environmental effects

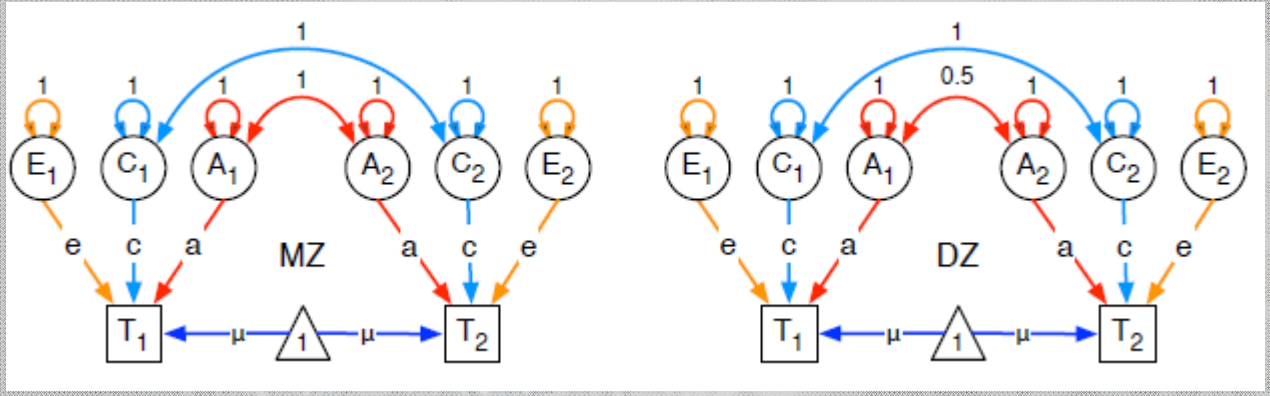
- Coefficient =1 for MZ and DZ pairs
- Equal environment assumption – for all the environmental influences THAT MATTER there is ON AVERAGE no differences in the degree of environmental sharing between MZ and DZ pairs

MZ		DZ	
$a^2+c^2+e^2$	a^2+c^2	$a^2+c^2+e^2$	$.5a^2+c^2$
a^2+c^2	$a^2+c^2+e^2$	$.5a^2+c^2$	$a^2+c^2+e^2$

Today we will run an ACE model

- Open RStudio
- faculty/sarah/2016/Tuesday1
- Copy everything

Today we will run an ACE model




```
# Create Matrices for Path Coefficients
pathA     <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE, values=svPa,
label="a11", lbound=lbPa, name="a" )
pathC     <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE, values=svPa,
label="c11", lbound=lbPa, name="c" )
pathE     <- mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE, values=svPe,
label="e11", lbound=lbPa, name="e" )
```

We are going to include the effect of age on the means

```
# Create Matrices for Covariates and linear Regression Coefficients
defAge    <- mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE,
labels=c("data.age"), name="Age" )
pathB     <- mxMatrix( type="Full", nrow=1, ncol=1, free=TRUE,
values=.01, label="b11", name="b" )

# Create Algebra for expected Mean Matrices
meanG     <- mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE,
values=svMe, labels="xbmi", name="meanG" )
expMean   <- mxAlgebra( expression= meanG + cbind(b%*%Age,b%*%Age),
name="expMean" )
```

Covariances

```
# Create Algebra for Variance Components
covA      <- mxAlgebra( expression=a %*% t(a), name="A" )
covC      <- mxAlgebra( expression=c %*% t(c), name="C" )
covE      <- mxAlgebra( expression=e %*% t(e), name="E" )
```

```
expCovMZ  <- mxAlgebra( expression= rbind( cbind(A+C+E, A+C),
                                           cbind(A+C, A+C+E)), name="expCovMZ" )
expCovDZ  <- mxAlgebra( expression= rbind( cbind(A+C+E, 0.5%x%A+C),
                                           cbind(0.5%x%A+C, A+C+E)), name="expCovDZ" )
```

MZ

$a^2+c^2+e^2$ a^2+c^2

a^2+c^2 $a^2+c^2+e^2$

DZ

$a^2+c^2+e^2$ $.5a^2+c^2$

$.5a^2+c^2$ $a^2+c^2+e^2$

Data	
dzData	351 obs. of 2 variables
mzData	569 obs. of 2 variables
twinData	3808 obs. of 12 variables
Values	

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

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

To fit a model to data, the differences between the observed covariance matrix and model-implied expected covariance matrix are minimized.
Objective functions are functions for which free parameter values are chosen such that the value of the objective function is minimized.
mxFitFunctionML() uses full-information maximum likelihood to provide maximum likelihood estimates of free parameters in the algebra defined by the covariance and means arguments.

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

```
# Create Algebra for Variance Components
rowVC     <- rep('VC',nv)
colVC     <- rep(c('A','C','E','SA','SC','SE'),each=nv)
estVC     <- mxAlgebra( expression=cbind(A,C,E,A/V,C/V,E/V),
             name="VC", dimnames=list(rowVC,colVC))

# Create Confidence Interval Objects
ciACE     <- mxCI( "VC[1,1:3]" )

# Build Model with Confidence Intervals
modelACE  <- mxModel( "oneACEca", pars, modelMZ, modelDZ, multi, estVC, ciACE )

# ------------------------------------------------------------------------------
# RUN MODEL

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

What to report

- Summary statistics
  - Usually from a simplified ‘saturated’ model
- Standardized estimates
  - Easier to conceptualise
    - ie 40% of the phenotypic variance vs a genetic effect of 2.84
    - Can easily be returned to original scale if summary statistics are provided

What to report

- Path coefficients
  - Very important in multivariate analyses
    - Gives a much clearer picture of the directionality of effects
- Variance components/proportion of variance explained
- Genetic correlations

How to report it?

- Path diagram
  - onyx (java based cross platform)
    - http://onyx.brandmaier.de/
  - Clasic MX GUI (complex to setup in mac)
    - http://www.vipbg.vcu.edu/vipbg/mxgui/
  - umx with graphviz (complex to setup in windows)
    - See Tim's talk from yesterday
  - omnigraffle (mac only)
    - https://www.omnigroup.com/omnigraffle

General Advice/Problem solving

- Scripting styles differ
- Check the sample description
- Learn to love the webpage
- Comments are your friends


Bus shelter on the road to
Sintra (Portugal)