



Mx Practical

TC18, 2005

Dorret Boomsma, Nick Martin,
Hermine H. Maes



Basic Genetic Epidemiology

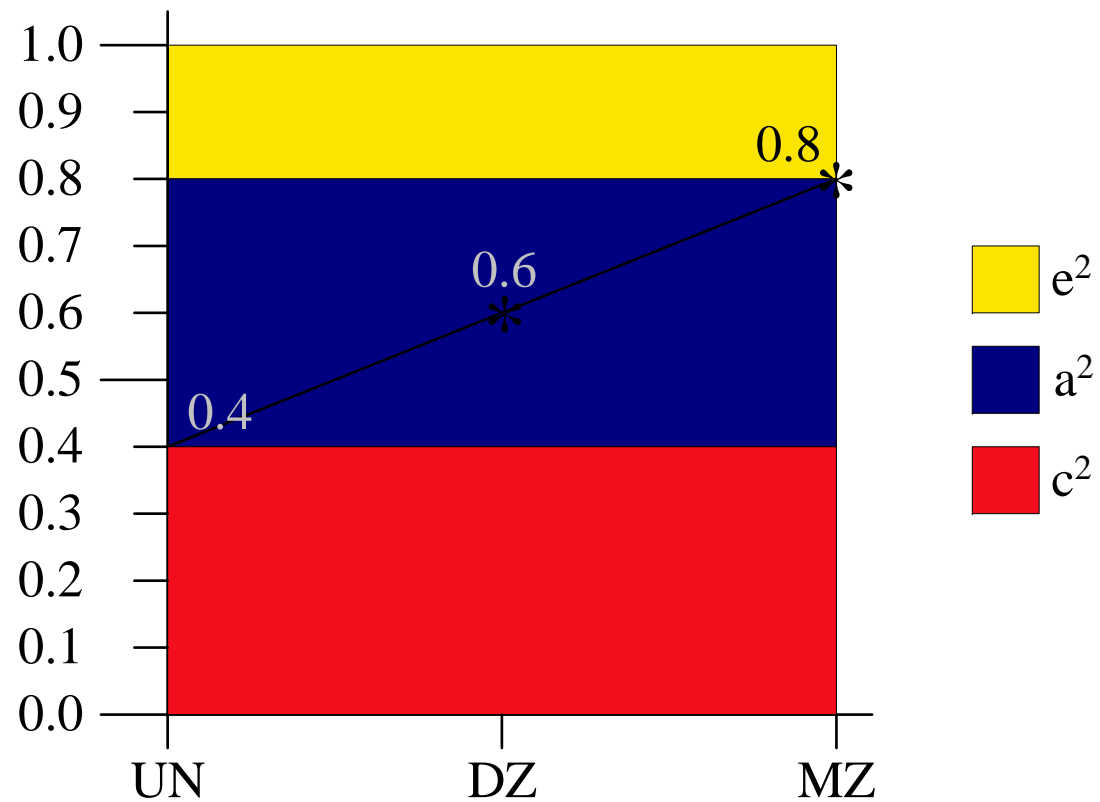
- Is the trait genetic?
 - Collect phenotypic data on large samples of MZ & DZ twins
 - Compare MZ & DZ correlations
 - Partition/ Quantify the variance in genetic and environmental components
 - Test significance of genetic variance



Practical Example

- Dataset: Dutch Adult Twins
- Cardiovascular Study: Fasting blood samples
- Variables: LDL, ApoB, ApoE(In)
- Age range: 35-60 years
- Sample Sizes: MZ: 91 pairs, DZ: 116 pairs

MZ & DZ correlations





Raw Dataset: DutchMZ.rec

```
3      1      44.5      2.35      1.01      12.92      1      44.5      3.21      1.31      12.21
3      1      40.5      2.70      0.91      13.77      1      40.5      2.93      1.04      14.53
3      1      44.5      3.68      1.10      11.41      1      44.5      3.71      1.23      12.69
3      1      37.5      2.34      0.88      6.84      1      37.5      1.73      0.80      6.84
3      1      55.5      2.46      1.17      10.19      1      55.5      3.88      1.46      12.92
1      0      40.5      4.86      1.61      13.14      0      40.5      5.03      1.67      14.35
3      1      52.5      5.74      1.99      16.27      1      52.5      5.96      1.67      17.66
.....
```

■ MZ twins

- Data NInput=11
- Rectangular File=DutchMZ.rec
- Labels zyg sex1 age1 ldl1 apob1 Inapoe1
sex2 age2 ldl2 apob2 Inapoe2



Univariate Genetic Analysis

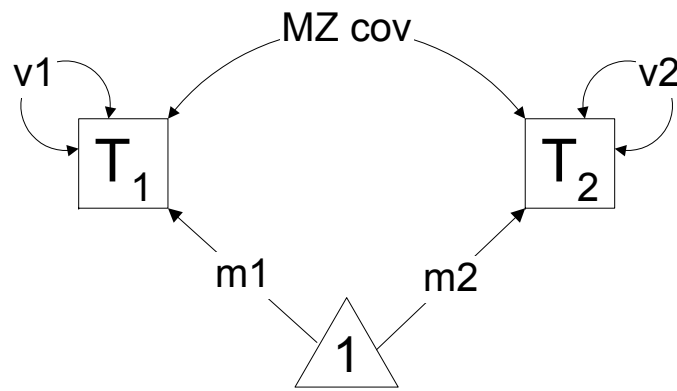
- Saturated Models

- Free variances, covariances $>$ correlations
- Free means

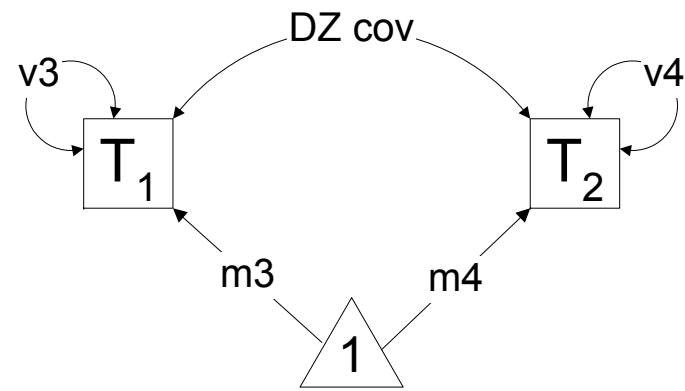
- Univariate Models

- Variances partitioned in a, c/d and e
- Free means (or not)

Free means, (co)variances



MZ twins



DZ twins

10 parameters

Correlation = covariance / square root of (variance1 * variance2)

Covariance = correlation * square root of (variance1 * variance2)



Mx Group Structure

- Title
 - Group type: data, calculation, constraint
 - [Read observed data, Labels, Select]
 - Matrices declaration
 - Begin Matrices; End Matrices;
 - [Specify numbers, parameters, etc.]
 - Algebra section and/or Model statement
 - Begin Algebra; End Algebra;
 - Means Covariances
 - [Options]
- End

! Estimate means and correlations

! Dutch Adult Twins: Lipid levels

```
■ #define nvar 1
■ #define nvarx2 2
■ #NGroups 2

■ G1: MZ twins
■   Data NInput=11
■   Rectangular File=DutchMZ.rec
■   Labels ..
■   Select ldl1 ldl2 ;
■   Begin Matrices;
■     M Full nvar nvarx2 Free
■     S Diag nvarx2 nvarx2 Free
■     R Stnd nvarx2 nvarx2 Free
■   End Matrices;
■   ! Starting values
■   Means M;
■   Covariance S*R*S';
■ End

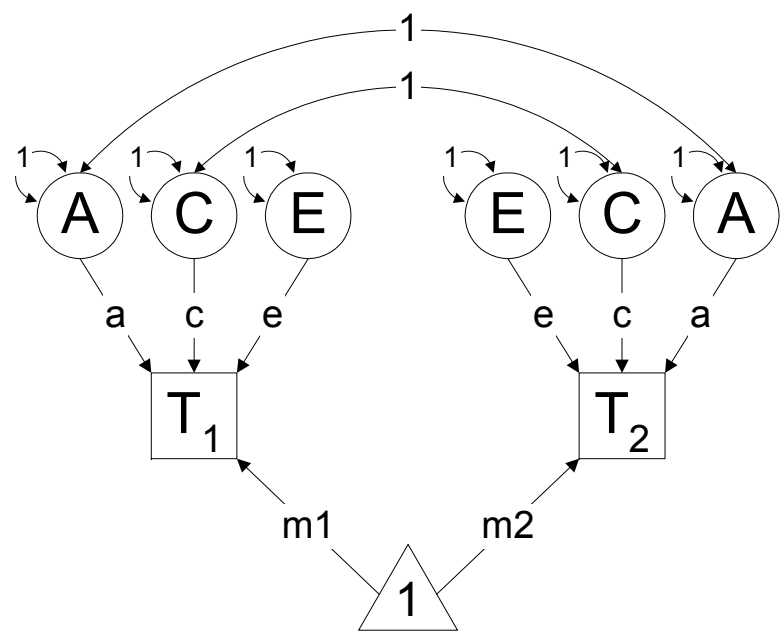
■ G2: DZ twins
■   Data NInput=18
■   Rectangular File=DutchDZ.rec
■   Labels ..
■   Select ldl1 ldl2 ;
■   Begin Matrices;
■     M Full nvar nvarx2 Free
■     S Diag nvarx2 nvarx2 Free
■     R Stnd nvarx2 nvarx2 Free
■   End Matrices;
■   ! Starting values
■   Means M;
■   Covariance S*R*S';
■ End
```

Correlations_MZ+DZ.mx

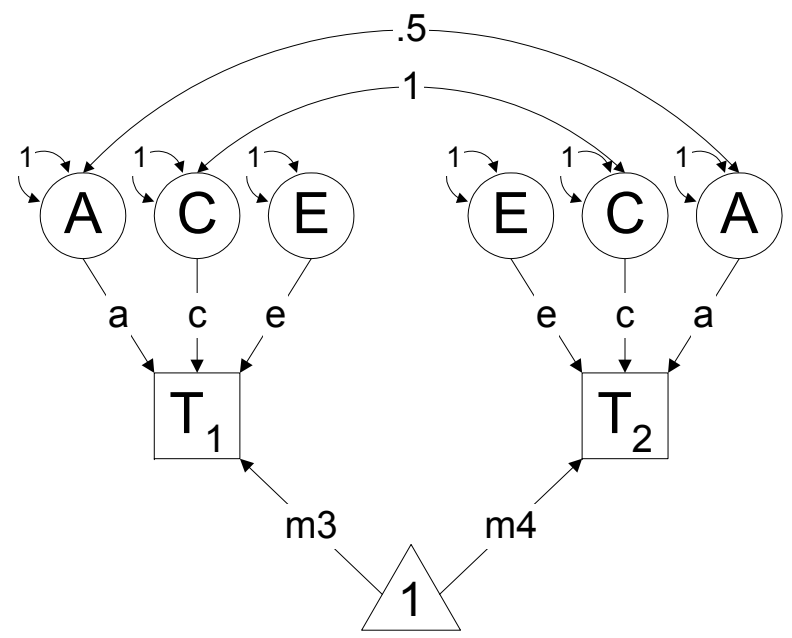
Correlations + Test MZ=DZcor

	MZ	DZ	Chi (df=1)	p
LDL	.78	.45	16.32	.000
ApoB	.79	.46	16.67	.000
lnApoE	.89	.51	33.31	.000

Means, ACE



MZ twins



DZ twins

7 parameters

Expected Covariances

Observed Cov	Variance Twin 1	Covariance T1T2
	Covariance T1T2	Variance Twin 2
MZ Expected Cov	$a^2+c^2+e^2+d^2$	$a^2+c^2+d^2$
	$a^2+c^2+d^2$	$a^2+c^2+e^2+d^2$
DZ Expected Cov	$a^2+c^2+e^2+d^2$	$.5a^2+c^2+.25d^2$
	$.5a^2+c^2+.25d^2$	$a^2+c^2+e^2+d^2$

! Estimate variance components - ACED model

! Dutch Adult Twins: Lipid levels

```
■ #define nvar 1
■ #define nvar2 2
■ #NGroups 4

■ Title 1: Model Parameters
■ Calculation
■ Begin Matrices;
■ X Lower nvar nvar Free ! a
■ Y Lower nvar nvar Free ! c
■ Z Lower nvar nvar Free ! e
■ W Lower nvar nvar ! d
■ H Full 1 1 ! 0.5
■ Q Full 1 1 ! 0.25
■ End Matrices;
■ Matrix H .5
■ Matrix Q .25

■ Label Row X add_gen
■ Label Row Y com_env
■ Label Row Z spec_env
■ Label Row W dom_gen
■ Begin Algebra;
■ A= X*X'; ! a^2
■ C= Y*Y'; ! c^2
■ E= Z*Z'; ! e^2
■ D= W*W'; ! d^2
■ End Algebra;
■ End
```

ACEmodel_MZ+DZ.mx

! Estimate variance components - ACED model

! Dutch Adult Twins: Lipid levels

- Title G2: MZ data
- Data NInput=11
- Rectangular File=DutchMZ.rec
- Labels ..
- Select ldl1 ldl2 ;
- Begin Matrices = Group 1;
- M Full 1 nvar2 Free
- End Matrices;
- Matrix M 4 4
- Means M;
- Covariance
- $$\begin{array}{c|c} A+C+E+D & A+C+D \\ \hline A+C+D & A+C+E+D \end{array}$$
- Option RSiduals
- End

- Title 3: DZ data
- Data NInput=18
- Rectangular File=DutchDZ.rec
- Labels ..
- Select ldl1 ldl2 ;
- Begin Matrices = Group 1;
- M Full 1 nvar2 Free
- End Matrices;
- Matrix M 4 4
- Means M;
- Covariance
- $$\begin{array}{c|c} A+C+E+D & H@A+C+Q@D \\ \hline H@A+C+Q@D & A+C+E+D \end{array}$$
- Option RSiduals
- End

ACEmodel_MZ+DZ.mx

! Estimate variance components - ACED model

! Dutch Adult Twins: Lipid levels

```
■ Title G4: Standardization
■ Calculation
■ Begin Matrices = Group 1;
■ End Matrices;
■ Start .6 all
■ Begin Algebra;
■ V=A+C+E+D;
■ P=A|C|E|D;
■ S=P@V~;
■ End Algebra;
■ Label Col P a^2 c^2 e^2 d^2
■ Label Col S a^2 c^2 e^2 d^2
■ !ACE model
■ ! Interval S 1 1 - S 1 4
■ Option NDecimals=4
■ Option Multiple Issat
■ End

■ Save ACE.mxs
■ ! Test significance of genetic ..
■ Drop X 1 1 1
■ End

■ Get ACE.mxs
■ ! Test significance of shared env
■ Drop Y 1 1 1
■ Exit
```

ACEmodel_MZ+DZ.mx




Chi-square tests and probs

	Significance of A (df=1)		Significance of C (df=1)	
	LDL	25.14	.000	0.30
ApoB	28.58	.000	0.16	.693
lnApoE	37.01	.000	1.97	.161



Variance Components ACE

	a^2	c^2	e^2	
LDL	.72	.08	.20	
ApoB	.75	.06	.19	
lnApoE	.67	.20	.13	



Using templates & headers in Mx

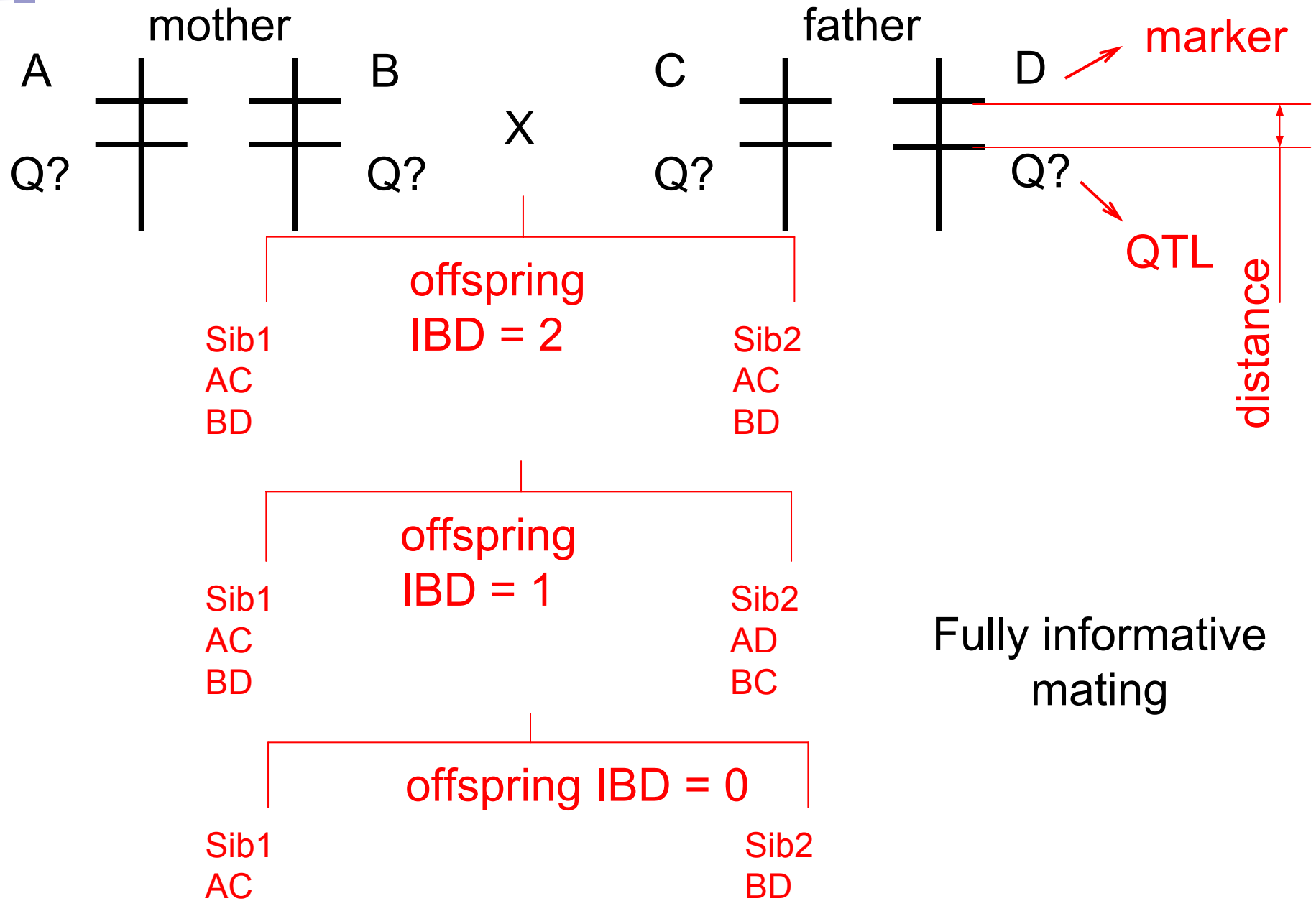
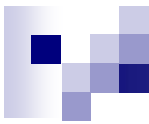
- Correlations_MZ+DZ_template.mx
 - #include DutchMZ.dat
 - Select
 - t1\$var
 - t2\$var ;

- Correlations_MZ+DZ_header.mx
 - #define \$var Idl
 - #include Correlations_MZ+DZ_template.mx



Linkage Analysis

- Where are the genes?
 - Collect genotypic data on large number of markers
 - Compare correlations by number of alleles identical by descent at a particular marker
 - Partition/ Quantify variance in genetic (QTL) and environmental components
 - Test significance of QTL effect

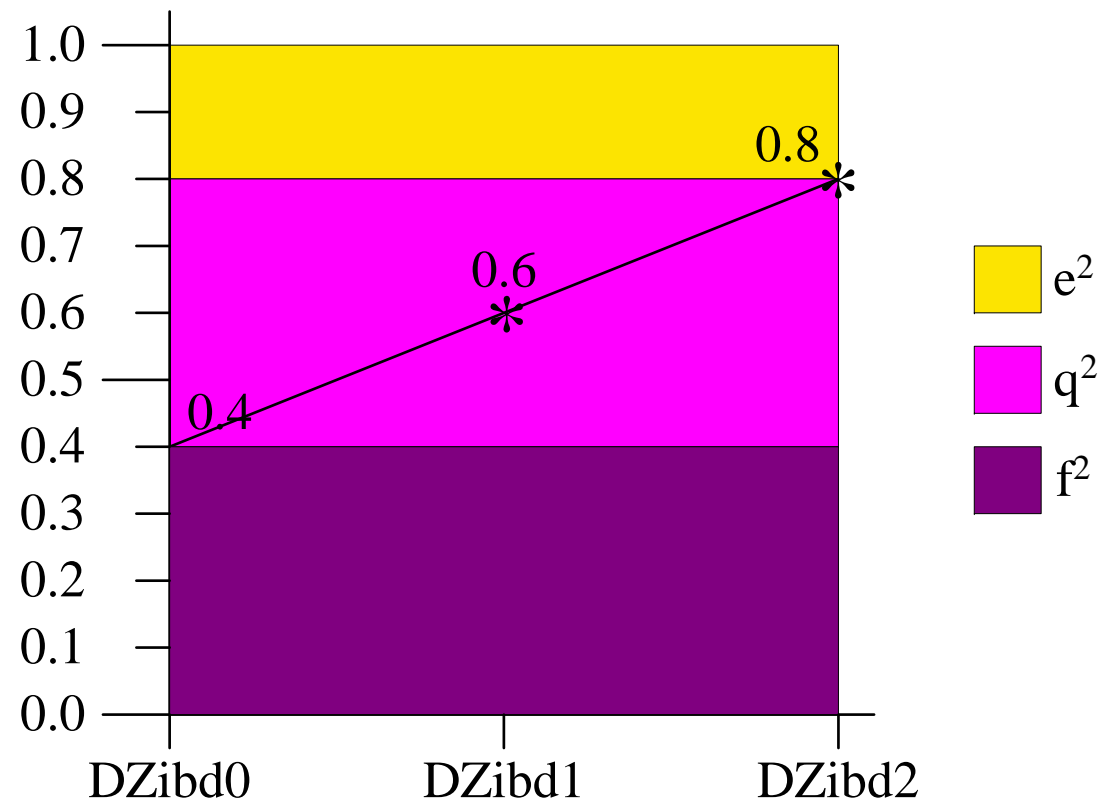


Identity by Descent (IBD) in sibs

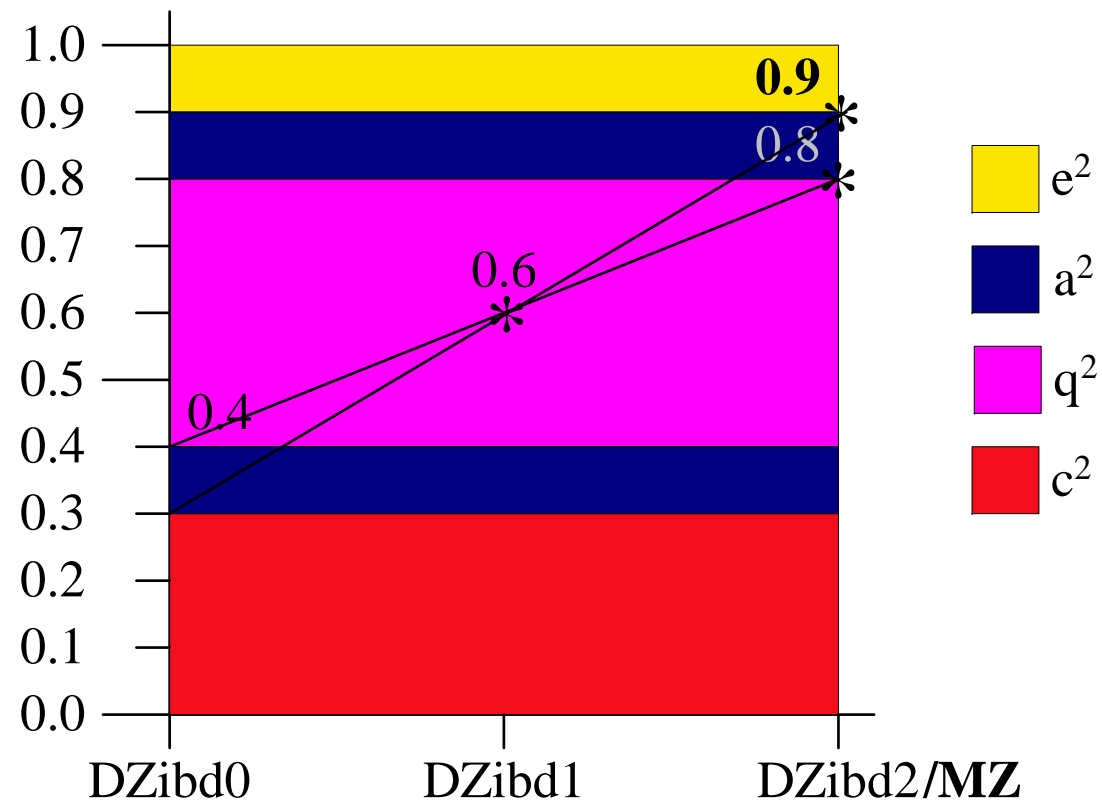
		Sib1			
		AC	AD	BC	BD
Sib2	AC	2	1	1	0
	AD	1	2	0	1
	BC	1	0	2	1
	BD	0	1	1	2

- Four parental marker alleles: A-B and C-D
- Two siblings can inherit 0, 1 or 2 alleles IBD
- IBD 0:1:2 = 25%:50%:25%
- Derivation of IBD probabilities at one marker (Haseman & Elston 1972)

DZ ibd0,1,2 correlations



DZ ibd0,1,2 & MZ correlations

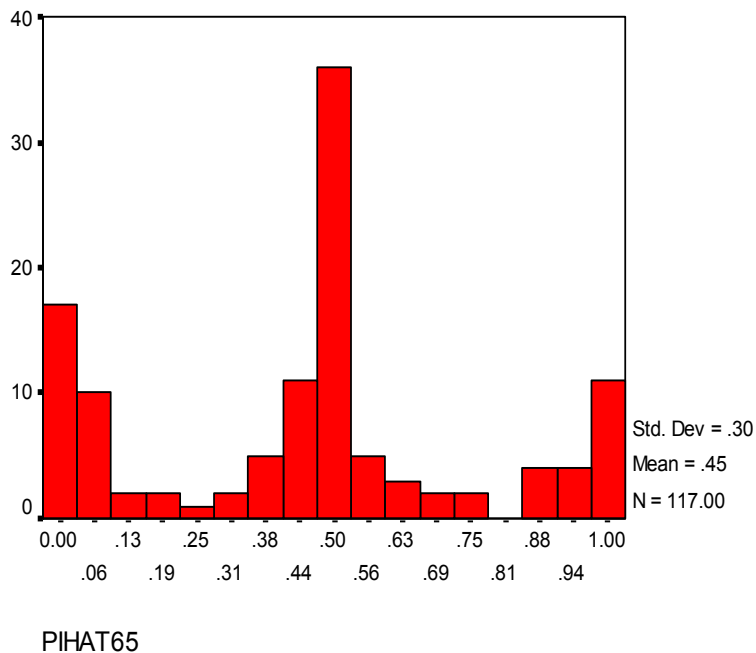




Raw Dataset: DutchDZ.rec

- DZ twins
 - Data NInput=18
 - Rectangular File= DutchDZ.rec
 - Labels zyg sex1 age1 med1 ldl1 apob1 lnapoe1 sex2 age2 med2 ldl2 apob2 lnapoe2 ibd0_65 ibd1_65 ibd2_65 pihat65 pi65cat
- position 65 on chromosome 19
- ibd0_65 ibd1_65 ibd2_65: probabilities that sibling pair is ibd 0, 1 or 2
- pihat65: pihat estimated as $\frac{1}{2}(\text{ibd1_65}) + (\text{ibd2_65})$
- pi65cat: sample divided according to $\pi < .25$, $\pi > .75$ or other

Distribution of pi-hat



- Adult Dutch DZ pairs: distribution of pi-hat (π) at 65 cM on chromosome 19
- $\pi = IBD/2$
 - $\pi < 0.25$: IBD=0 group
 - $\pi > 0.75$: IBD=2 group
 - others: IBD=1 group
 - pi65cat= (0,1,2)



- Can resemblance (e.g. correlations, covariances) between sib pairs, or DZ twins, be modeled as a function of DNA marker sharing at a particular chromosomal location?



Compare correlations by IBD

- DZ pairs (3 groups according to IBD) only
 - Estimate correlations as function of IBD (pi65cat)
 - Test if correlations are equal



Add MZ twins

- DZ + MZ pairs
 - Estimate correlations as function of IBD + zygosity
 - Test if DZ_{ibd2} correlation is equal to MZ correlation



Correlations

	DZibd2	DZibd1	DZibd0	MZ
LDL	.81	.49	-.21	.78
ApoB	.64	.50	.02	.79
lnApoE	.83	.55	.14	.89



Tests

-
- Option Multiple Issat
- End

- Save lipidcor.mxs
- ! Test for linkage
- ! Set 3 DZ IBD correlations equal
- Equate R 1 2 1 R 2 2 1 R 3 2 1
- End

- Get lipidcor.mxs
- ! Test for residual polygenic variance
- ! Set DZ IBD2 correlation equal to MZ correlation
- Equate R 1 2 1 R 4 2 1
- Exit

Chi-square tests and probs

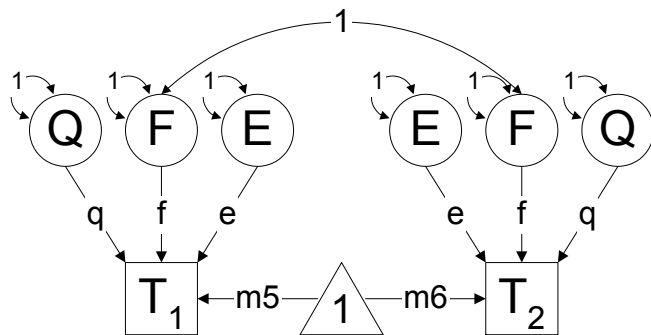
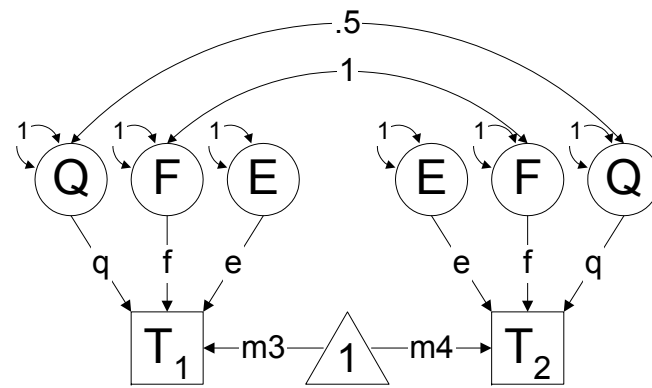
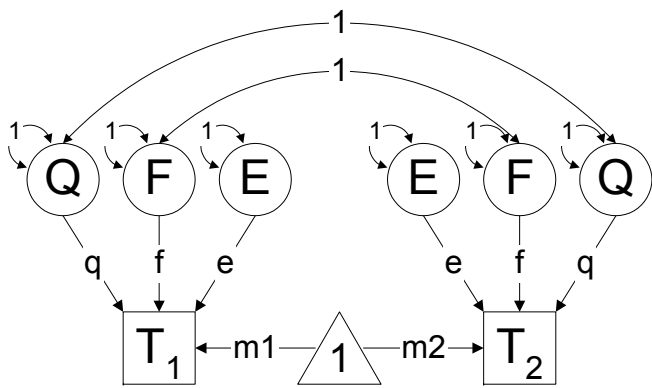
	All DZ equal (df=2)		DZibd2 = MZ (df=1)	
LDL	21.77	.000	0.09	.757
ApoB	7.98	.019	1.53	.216
lnApoE	12.45	.002	0.58	.448



Compare correlations

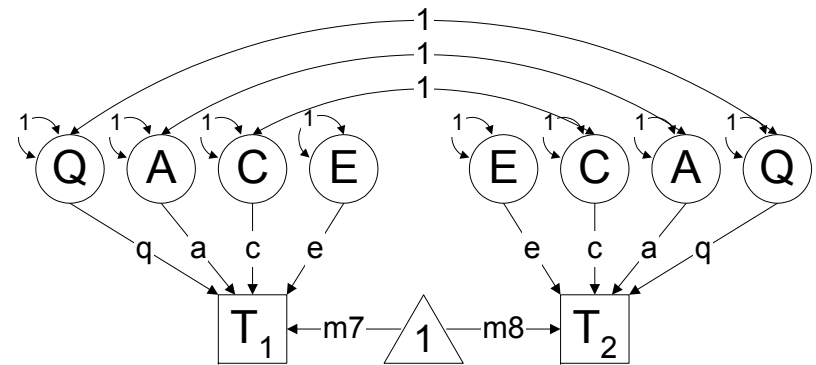
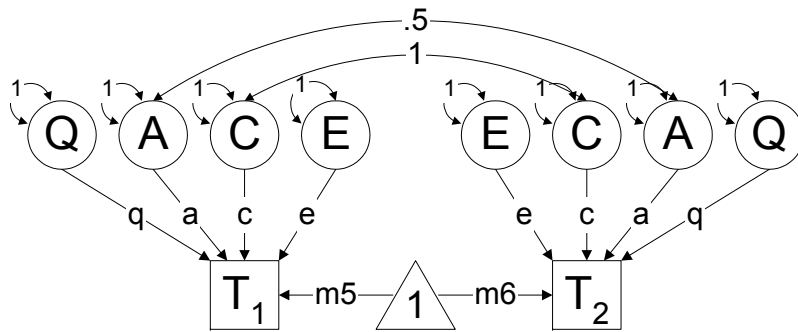
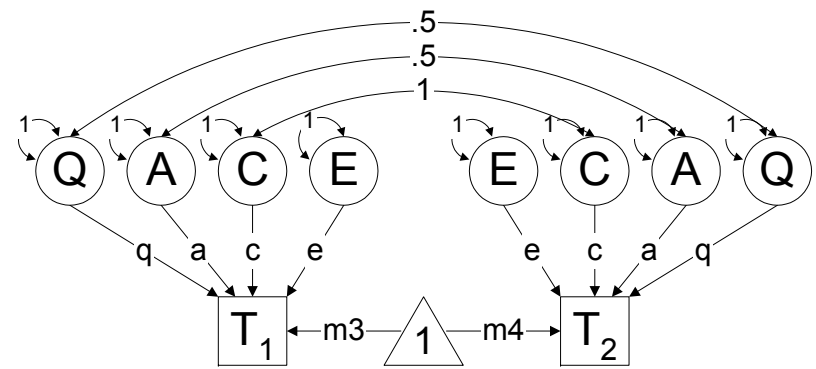
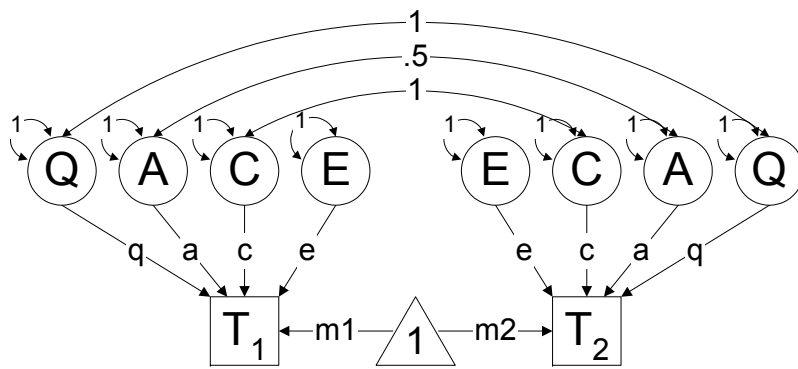
- DZ pairs (3 groups according to IBD) only
 - Estimate correlations as function of IBD
 - Test if correlations are equal
 - Correlations_DZibd.mx
- DZ + MZ pairs
 - Estimate correlations as function of IBD + zyg
 - Test if DZibd2 correlation is equal to MZ cor
 - Correlations_DZibd+MZ.mx

DZ by IBD status



- Variance = $Q + F + E$
- Covariance = $\pi Q + F + E$

DZ by IBD status + MZ





Partition Variance

- DZ pairs (3 groups according to IBD) only
 - Estimate FEQ
 - Test if QTL effect is significant

Covariance Statements

- G2: DZ IBD2 twins
- Matrix K 1
- Covariance
- $F+Q+E \mid F+K@Q \text{ _}$
- $F+K@Q \mid F+Q+E;$

- G3: DZ IBD1 twins
- Matrix K .5
- Covariance
- $F+Q+E \mid F+K@Q \text{ _}$
- $F+K@Q \mid F+Q+E;$

- G4: DZ IBD0 twins
- Covariance
- $F+Q+E \mid F \text{ _}$
- $F \mid F+Q+E;$



Partition Variance

- DZ + MZ pairs
 - Estimate ACEQ
 - Test if QTL estimate/significance is different

Covariance Statements +MZ

- G2: DZ IBD2 twins
 - Matrix K 1
 - Covariance
 - $A+C+Q+E \mid H@A+C+K@Q _$
 - $H@A+C+K@Q \mid A+C+Q+E;$

- G3: DZ IBD1 twins
 - Matrix K .5
 - Covariance
 - $A+C+Q+E \mid H@A+C+K@Q _$
 - $H@A+C+K@Q \mid A+C+Q+E;$

- G4: DZ IBD0 twins
 - Covariance
 - $A+C+Q+E \mid H@A+C_$
 - $H@A+C \mid A+C+Q+E;$

- G5: MZ twins
 - Covariance
 - $A+C+Q+E \mid A+C+Q _$
 - $A+C+Q \mid A+C+Q+E;$



Chi-square Tests for QTL

	DZ pairs (df=1)		DZ+MZ pairs (df=1)	
LDL	12.25	.000	12.56	.000
ApoB	1.95	.163	2.13	.145
InApoE	12.45	.000	12.29	.000



Variance Components FEQ

		f^2	e^2	q^2
LDL		.00 (.00-.32)	.23 (.13-.40)	.77 (.36-.87)
ApoB		.27 (.00-.54)	.41 (.24-.66)	.32 (.00-.73)
InApoE		.19 (.00-.43)	.16 (.09-.32)	.65 (.33-.90)



Variance Components ACEQ

	a^2	c^2	e^2	q^2
LDL	.04 (.00-.39)	.00 (.00-.27)	.21 (.15-.29)	.75 (.37-.84)
ApoB	.46 (.11-.84)	.02 (.00-.29)	.19 (.24-.27)	.33 (.00-.67)
InApoE	.02 (.00-.33)	.22 (.00-.45)	.13 (.10-.18)	.63 (.32-.89)



Partition Variance

- DZ pairs (3 groups according to IBD) only
 - Estimate QFE
 - Test if QTL effect is significant
 - FEQmodel_DZibd.mx
- DZ + MZ pairs
 - Estimate
 - Test if QTL estimate/significance is different
 - ACEQmodel_DZibd+MZ.mx