# Mx Practical

TC20, 2007

Hermine H. Maes

Nick Martin, Dorret Boomsma

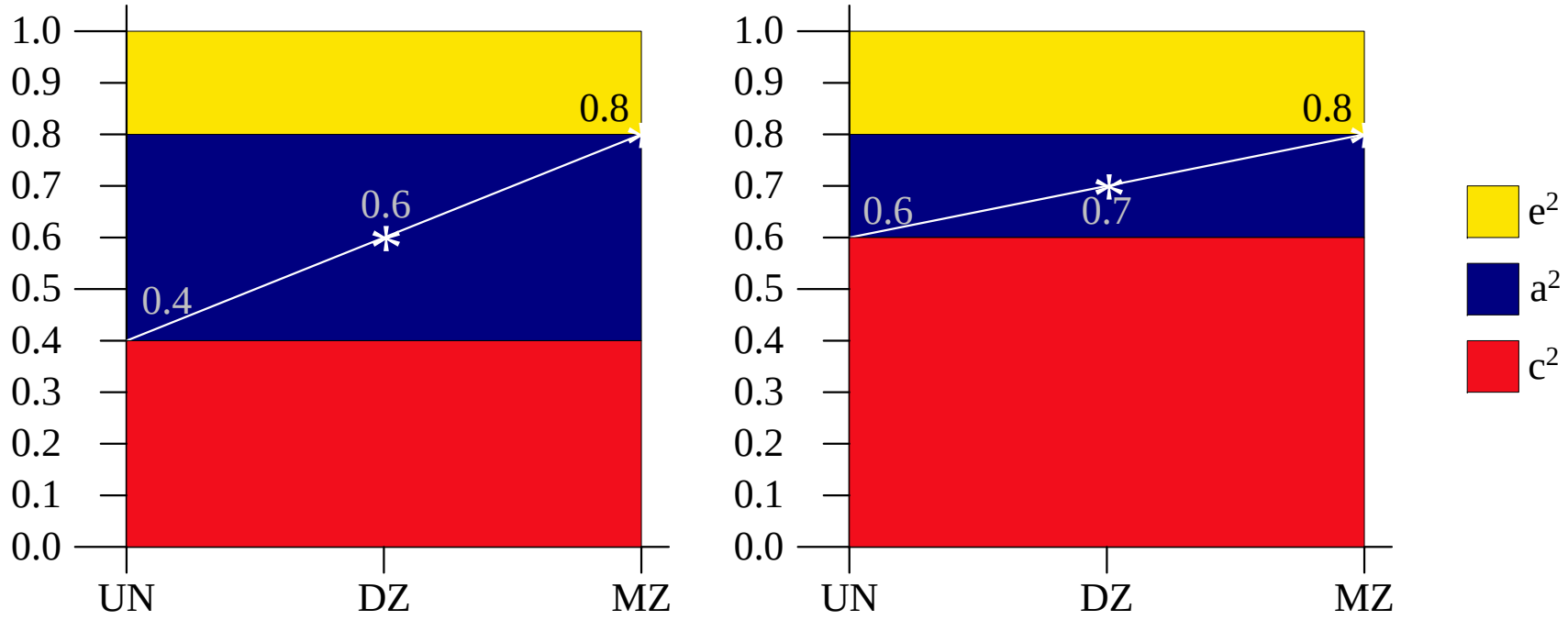
# Outline

- Intro to Genetic Epidemiology
- Progression to Linkage via Path Models
- Partitioned Twin Analyses
- Linkage using Pi-Hat
- Run Linkage in Mx

# 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

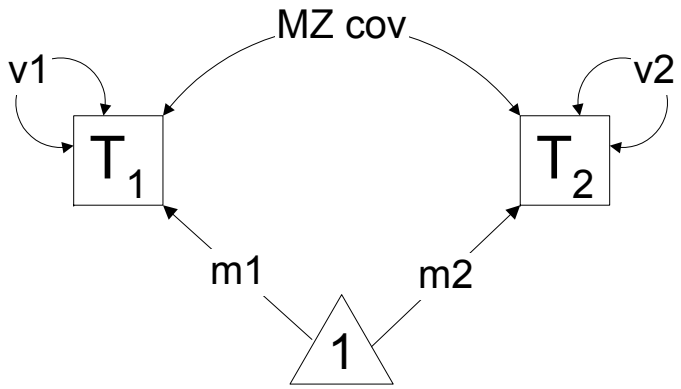
# MZ & DZ correlations



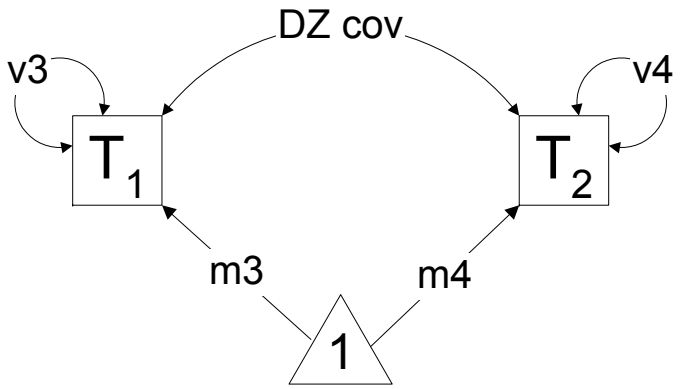
# Univariate Genetic Analysis

- 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



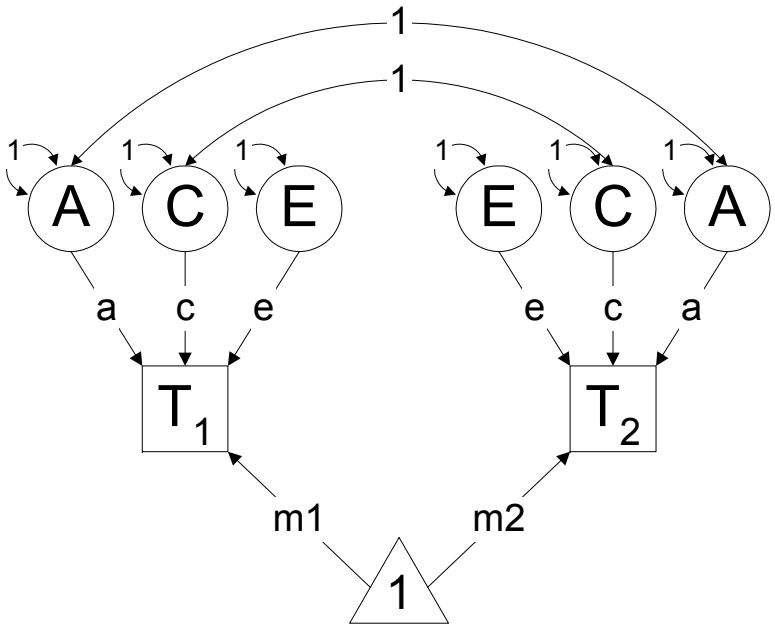

DZ twins

10 parameters

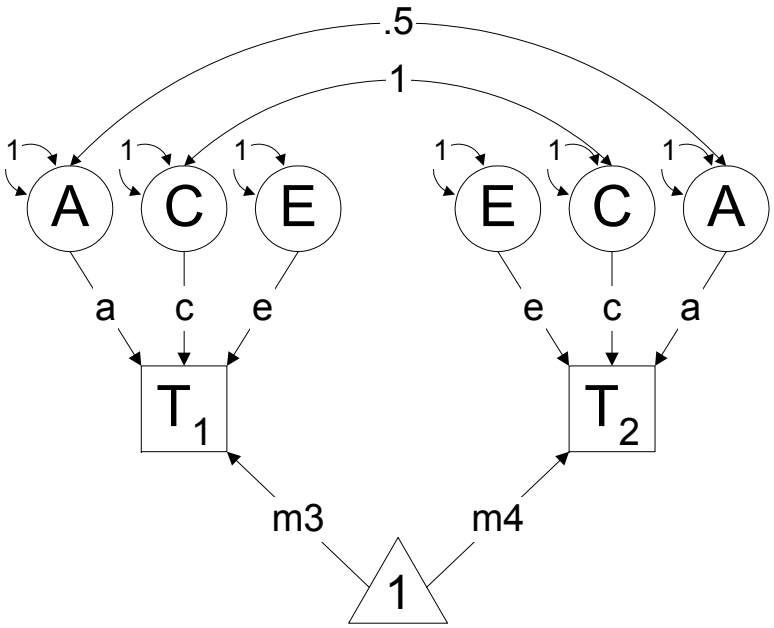
Correlation = covariance / square root of (variance1 * variance2)

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

# Means, ACE




MZ twins




DZ twins

7 parameters

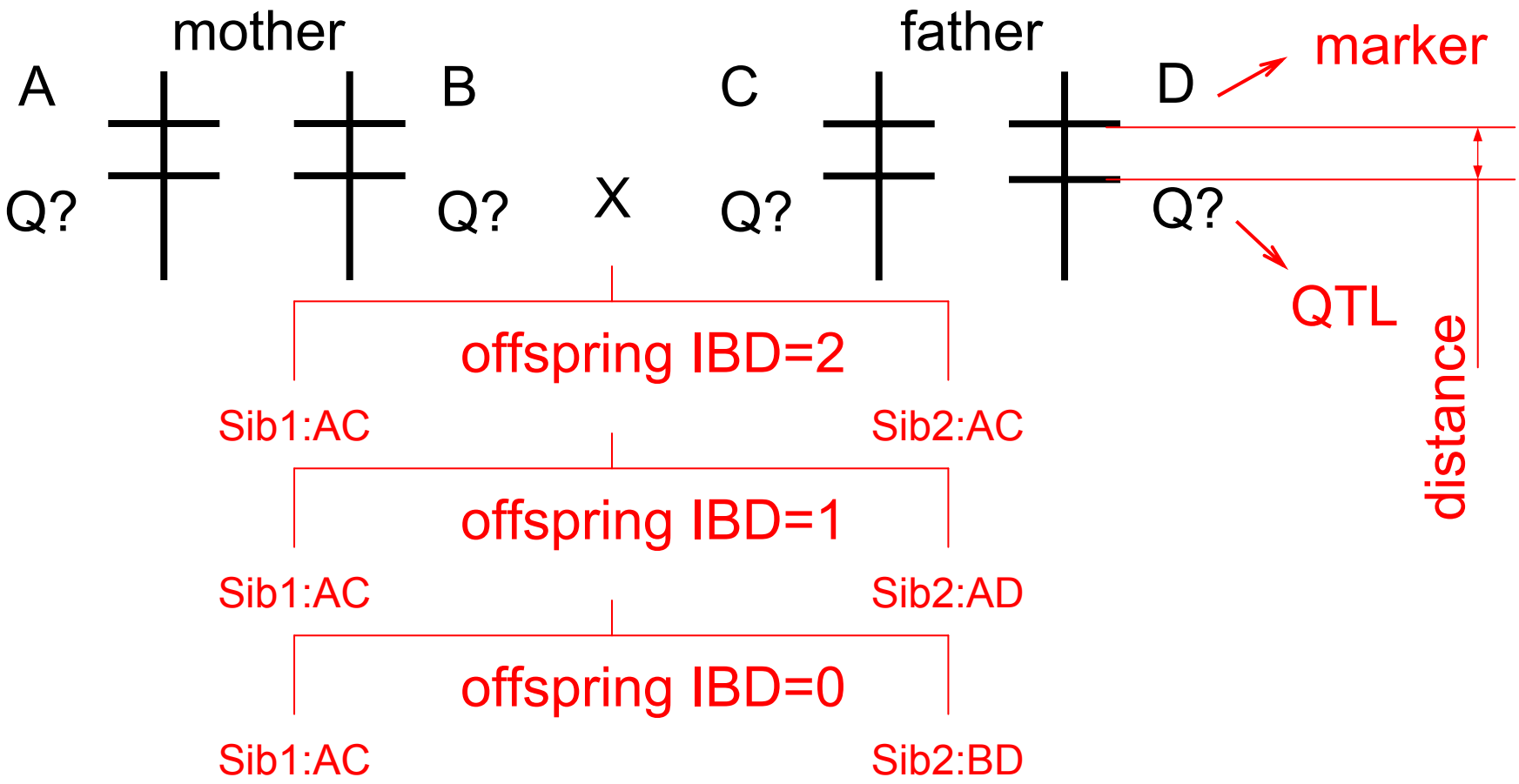
# Expected Covariances

| 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 | $a^2+c^2+e^2+d^2$ | $.5a^2+c^2+.25d^2$ |
| | $.5a^2+c^2+.25d^2$ | $a^2+c^2+e^2+d^2$ |

# 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

# Fully Informative Mating



# Identity by Descent (IBD) in sibs

| 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 |

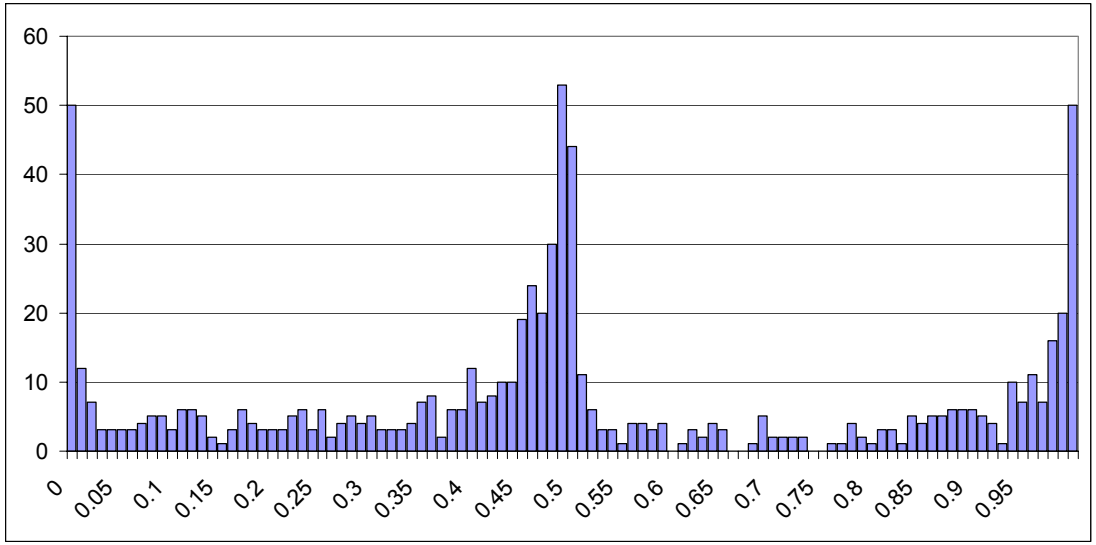
- Four parental marker alleles: A-B and C-D
- Two siblings can inherit 0, 1 or 2 alleles IBD
- IBD 0:1:2 = 25%:50%:25%
- Derivation of IBD probabilities at one marker (Haseman & Elston 1972

# Average IBD Sharing: Pi-hat

- Sharing at a locus can be quantified by the estimated proportion of alleles IBD
- Pi-hat = 0 x p(IBD=0) +
  .5 x p(IBD=1) +
  1 x p(IBD=2)

= p(IBD=2) + .5 x p(IBD=1)

# Distribution of pi-hat

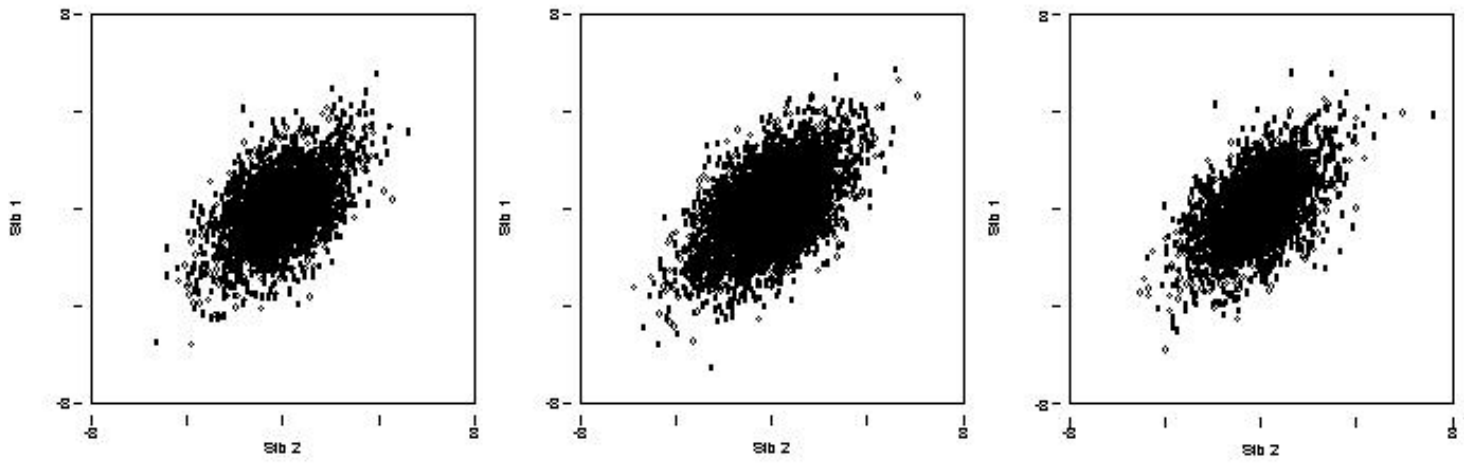



- DZ pairs: distribution of pi-hat (π) at particular cM on chromosome 2
- π<0.25: IBD=0 group  π>0.75: IBD=2 group
  others: IBD=1 group
- picat= (0,1,2)

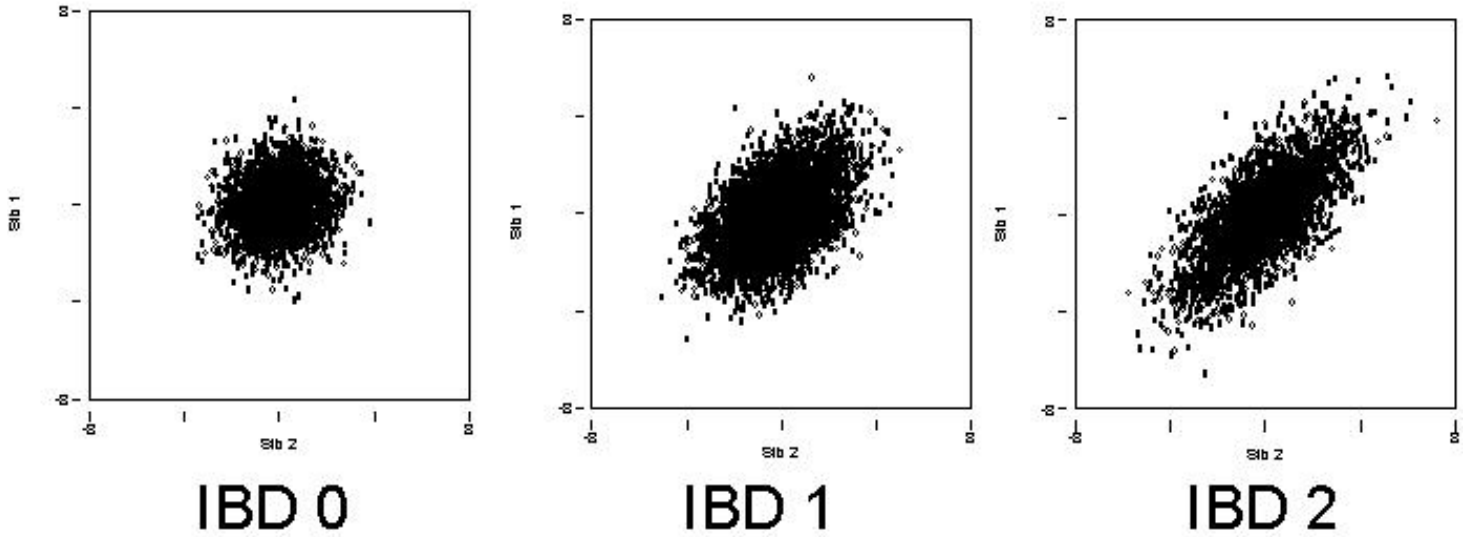
# Incorporating IBD

- Can resemblance (e.g. correlations, covariances) between sib pairs, or DZ twins, be modeled as a function of DNA marker sharing (IBD) at a particular chromosomal location?
  - Estimate covariance by IBD state
  - Impose genetic model and estimate model parameters

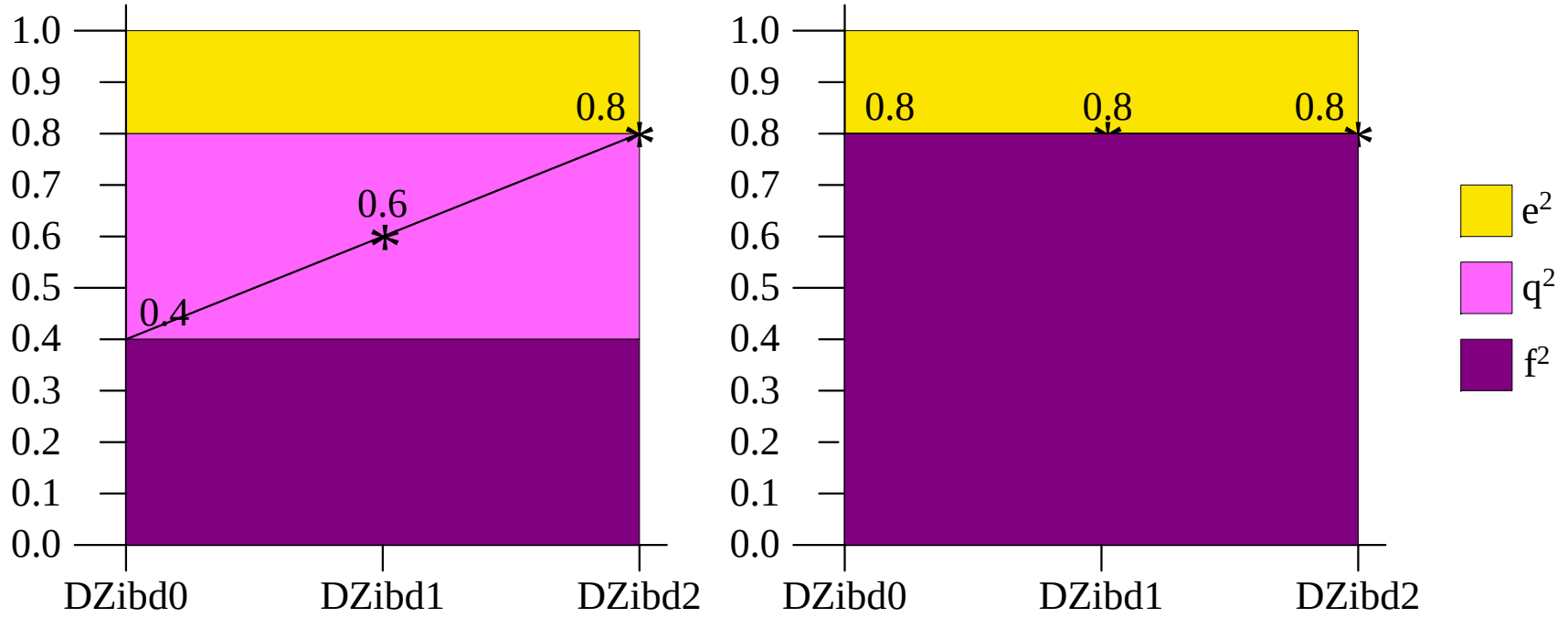
# No linkage



# Under linkage



# DZ ibd0,1,2 correlations



# Compare correlations by IBD

- DZ pairs (3 groups according to IBD) only
  - Estimate correlations as function of IBD (pi40cat)
  - Test if correlations are equal

# Typical Application

- Trait where genetic component is likely
  - Collect sample of relatives
  - Calculate IBD along chromosome
  - Test whether IBD sharing explains part of covariance between relatives

# Real data Example

- Gene Finding for intelligence
  - Intelligence is highly heritable (60-80%)
  - Actual genes not yet identified
- Two strategies:
  - Whole genome linkage analysis
  - Genetic association analysis

# Brisbane Twin and Family Study of Cognition

Memory
Attention
Problem
Solving

Queensland Institute of Medical Research and University of Queensland, Australia

# Publications

**Genome-wide Scan of IQ Finds Significant Linkage to a Quantitative Trait Locus on 2q**

*Behavior Genetics, Vol. 36, No. 1, January 2006*

M. Luciano,[1,2] M. J. Wright,[1] D. L. Duffy,[1] M. A. Wainwright,[1] G. Zhu,[1] D. M. Evans,[1] G. M. Geffen,[1] G. W. Montgomery,[1] and N. G. Martin[1]

**A Genomewide Scan for Intelligence Identifies Quantitative Trait Loci on 2q and 6p**

*Am. J. Hum. Genet. 77:318–326, 2005*

Danielle Posthuma,[1] Michelle Luciano,[2] Eco J. C. de Geus,[1] Margie J. Wright,[2] P. Eline Slagboom,[3] Grant W. Montgomery,[2] Dorret I. Boomsma,[1] and Nicholas G. Martin[2]

# Example Dataset

- 710 sib-pairs
- Performance IQ Data
- Chromosome 2
- 59 micro-satellite markers

# 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

# Raw Dataset

**piqDZ.rec**

80020 11 12 118 112 0.43647 0.55668 0.00685 0.28519 1
80030 12 11 121 127 0.0813 0.9187 0 0.45935 1
80033 11 12 113 123 0.03396 0.96604 0 0.48302 1
80040 12 11 125 94 0.00711 0.99289 0 0.496445 1
80090 11 12 87 80 0.02613 0.97387 0 0.486935 1
….

DZ twins

- `Data NInput=10`
- `  Rectangular File=piqDZ.rec`
- `  Labels fam id1 id2 piq1 piq2 ibd0mnr ibd1mnr ibd2mnr pihat picat`

- position ? on chromosome 2
- ibd0mnr ibd1mnr ibd2mnr: probabilities that sibling pair is ibd 0, 1 or 2
- pihat: pihat estimated as ½(ibd1mnr) + (ibd2mnr)
- picat: sample divided according to π<.25, π>.75 or other

```
Estimate Means and Correlations
#define nvar 1
#define nvarx2 2
#NGroups 3

G1: DZ IBD2 twins
 Data NInput=10
  Rectangular File=piqDZ.rec
  Labels fam ....
  Select if picat =2;
  Select piq1 piq2 ;
 Begin Matrices;
  M Full nvar nvarx2 Free       ! means
  S Diag nvarx2 nvarx2 Free     ! standard deviations
  R Stnd nvarx2 nvarx2 Free     ! correlations
 End Matrices;
  Matrix M 110 110   …          ! starting values
 Means M;
 Covariance S*R*S';
 Option RSiduals Mx%P=piq.il
End
```

Correlations_DZibd.mx

.. .. IBD1 & IBD0 groups

# Practical Correlations

- Mx script: Correlations_DZibd.mx

- Add groups for IBD1 and IBD0
- Test equality of correlations

faculty\hmaes\a20\maes\MxLinkage\

# Correlations

| | DZibd2 | DZibd1 | DZibd0 |
|---|---|---|---|
| piq | .60 | .27 | .15 |

# Test for Linkage

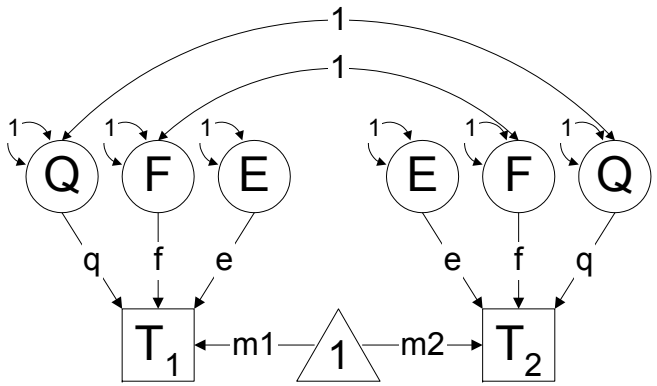
- Last Group of previous job
- ....
- Option Multiple Issat
- End

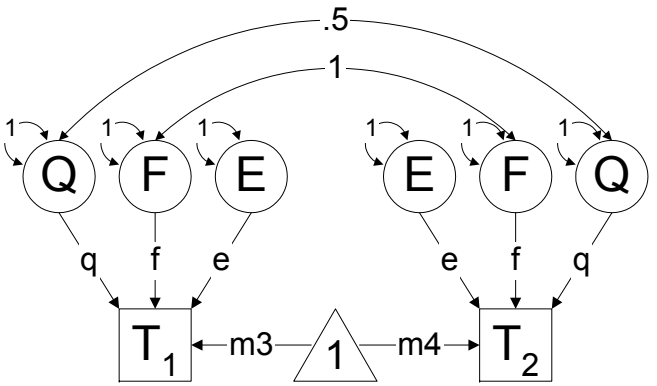
- Save piqcor.mxs
- ! Test for linkage
- ! Set 3 DZ IBD correlations equal
- Equate R 1 2 1 R 2 2 1 R 3 2 1
- End

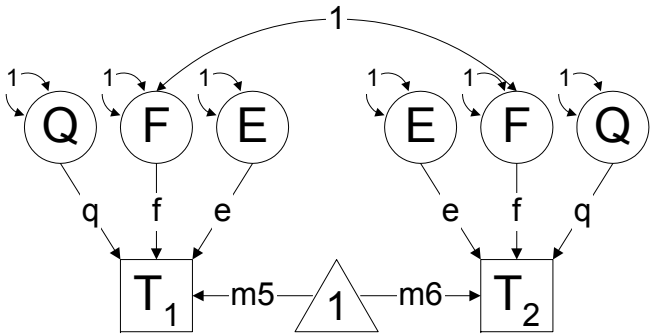
# Chi-square test and probability

| | All DZ equal | | |
|---|---|---|---|
| | $\chi^2$ | df | p |
| piq | 13.32 | 2 | .001 |

# DZ by IBD status




- Variance = Q + F + E
- Covariance = πQ + F + E

# Partition Variance

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

```
Estimate Variance Components: FEQ model
#define nvar 1
#define nvarx2 2
#NGroups 5

G1: Model Parameters
Calculation
 Begin Matrices;
  X Lower nvar nvar Free     ! residual familial paths
  Z Lower nvar nvar Free     ! unique environment paths
  L Lower nvar nvar Free     ! QTL path coefficients
  H Full 1 1
 End Matrices;
  Matrix H .5
  Start 5 All
 Begin Algebra;
  F=X*X';                    ! residual familial VC
  E=Z*Z';                    ! nonshared environment VC
  Q=L*L';                    ! QTL variance components
 End Algebra;
 Option Rsiduals
End
```

FEQmodel_DZ.mx

```
G2: DZ IBD2 twins
 Data NInput=10
  Rectangular File=piqDZ.rec
  Labels fam id1 id2 piq1 piq2 ibd0mnr ibd1mnr ibd2mnr
    pihat picat
  Select if picat =2;
  Select piq1 piq2 ;
 Begin Matrices = Group 1;
  M Full nvar nvarx2 Free
  K Full 1 1                    ! correlation QTL effects
 End Matrices;
  Matrix M 110 110
  Matrix K 1
 Means M;
 Covariance
    F+Q+E | F+K@Q _
    F+K@Q | F+Q+E;
End
```

FEQmodel_DZibd.mx

```
G3: DZ IBD1 twins
 Data NInput=10
  Rectangular File=piqDZ.rec
  Labels fam id1 id2 piq1 piq2 ibd0mnr ibd1mnr ibd2mnr
    pihat picat
  Select if picat =1;
  Select piq1 piq2 ;
 Begin Matrices = Group 1;
  M Full nvar nvarx2 Free
  K Full 1 1                    ! correlation QTL effects
 End Matrices;
  Matrix M 110 110
  Matrix K .5
 Means M;
 Covariance
    F+Q+E | F+K@Q _
    F+K@Q | F+Q+E;
End
```

FEQmodel_DZibd.mx

```
G4: DZ IBD0 twins
 Data NInput=10
  Rectangular File=piqDZ.rec
  Labels fam id1 id2 piq1 piq2 ibd0mnr ibd1mnr ibd2mnr
    pihat picat
  Select if picat =0;
  Select piq1 piq2 ;
 Begin Matrices = Group 1;
  M Full nvar nvarx2 Free
  K Full 1 1                    ! correlation QTL effects
 End Matrices;
  Matrix M 110 110
  Matrix K 1
 Means M;
 Covariance
    F+Q+E | F     _
    F     | F+Q+E;
End
```

FEQmodel_DZibd.mx

```
G5: Standardization
 Calculation
 Begin Matrices = Group 1;
 Begin Algebra;
  V=F+E+Q;                    ! total variance
  P=F|E|Q;                    ! concatenate estimates
  S=P@V~;                     ! standardized estimates
 End Algebra;
  Label Col P f^2 e^2 q^2
  Label Col S f^2 e^2 q^2
!FEQ model
 Interval S 1 1 - S 1 3
 Option Rsiduals NDecimals=4
 Option Multiple Issat
End

! Test for QTL
 Drop L 1 1 1
End
```

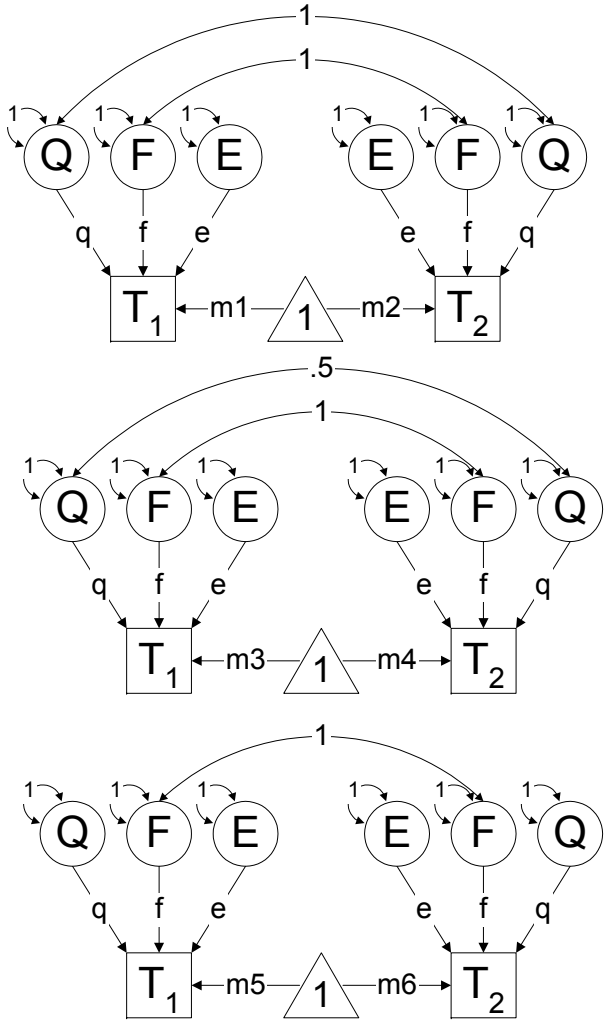
FEQmodel_DZibd.mx

# Covariance Statements

- G2: DZ IBD2 twins
- Matrix K 1
- Covariance
- F+Q+E | F+K@Q _
- F+K@Q | F+Q+E;

- G3: DZ IBD1 twins
- Matrix K .5
- Covariance
- F+Q+E | F+K@Q _
- F+K@Q | F+Q+E;

- G4: DZ IBD0 twins
- Covariance
- F+Q+E | F_
- F | F+Q+E;



# Chi-square test for QTL

| | All DZ pairs | | |
|---|---|---|---|
| | $\chi^2$ | df | p |
| piq | 13.07 | 1 | .000 |

# Variance Components FEQ

| | $f^2$ | $e^2$ | $q^2$ |
|---|---|---|---|
| piq | .10 (.00-.27) | .43 (.32-.58) | .46 (.22-.67) |

| | $a^2$ | $e^2$ | $q^2$ |
|---|---|---|---|
| piq | .21 (.00-.54) | .33 (.14-.52) | .47 (.22-.67) |

# Genome Scan

- Run multiple linkage jobs
  - Run ‘at the Marker’
  - Run ‘over a Grid’
    - Every 1/2/5/ cM?
- Pre-prepare your data files
  - One per chromosome or one per marker

# Merlin Output (merlin.ibd)

```
FAMILY ID1 ID2 MARKER P0 P1 P2
80020 3 3 2.113  0.0 0.0 1.0
80020 4 3 2.113  1.0 0.0 0.0
80020 4 4 2.113  0.0 0.0 1.0
80020 12 3 2.113  0.0 1.0 0.0
80020 12 4 2.113  0.0 1.0 0.0
80020 12 12 2.113  0.0 0.0 1.0
80020 11 3 2.113  0.0 1.0 0.0
80020 11 4 2.113  0.0 1.0 0.0
80020 11 12 2.113  0.32147 0.67853 0.00000
80020 11 11 2.113  0.0 0.0 1.0
80020 3 3 12.572  0.0 0.0 1.0
80020 4 3 12.572  1.0 0.0 0.0
80020 4 4 12.572  0.0 0.0 1.0
80020 12 3 12.572  0.0 1.0 0.0
80020 12 4 12.572  0.0 1.0 0.0
80020 12 12 12.572  0.0 0.0 1.0
80020 11 3 12.572  0.0 1.0 0.0
80020 11 4 12.572  0.0 1.0 0.0
80020 11 12 12.572  0.70372 0.29628 0.00000
```

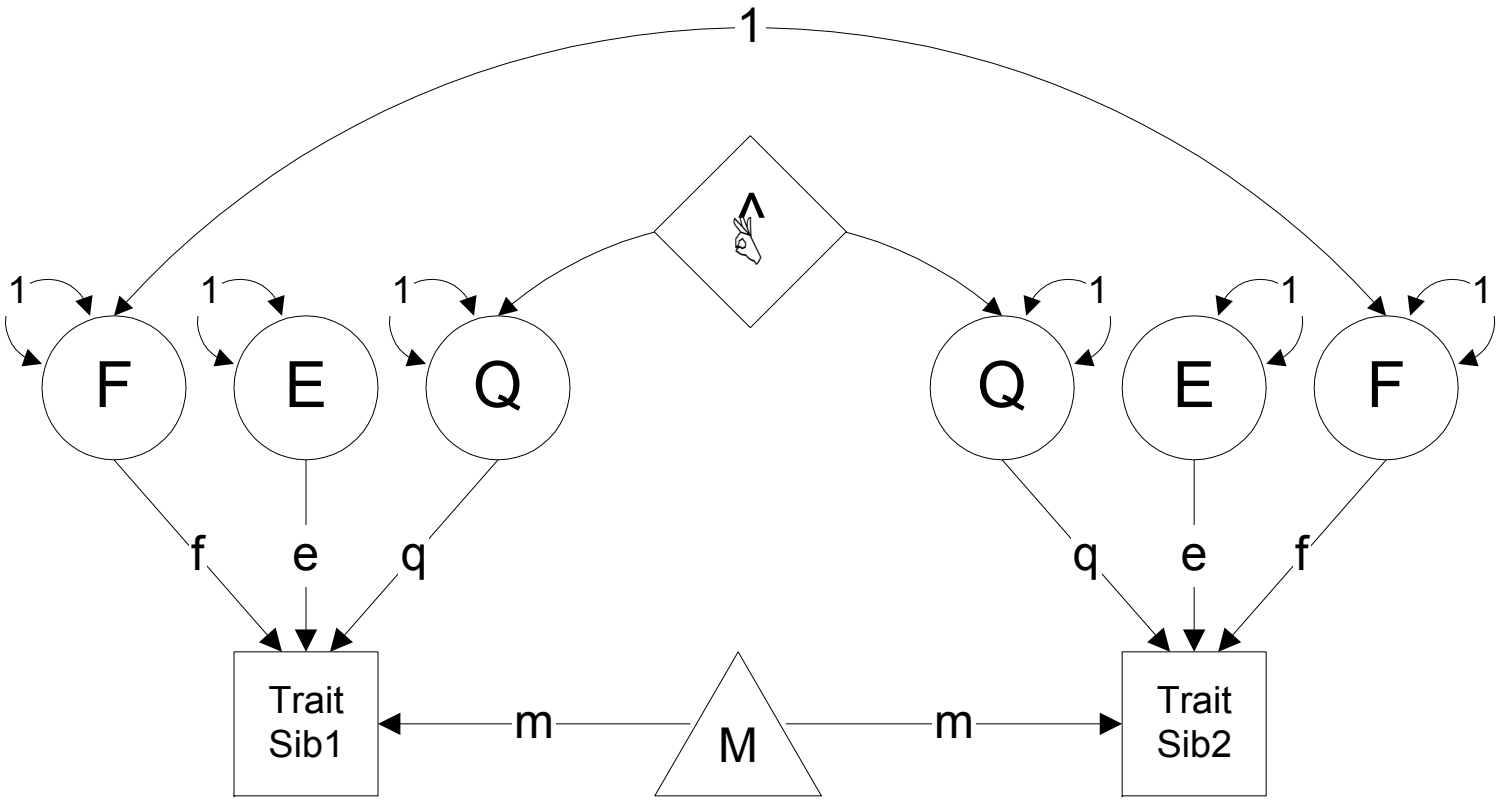
# Mx Input (piqibd.rec)

- 80020 11 12 118 112 0.32147 0.67853 0 0.70372 0.29628 0 1 0 0 0.99529 0.00471 0 1 0 0 0.27173 0.72827 0 0.25302 0.74171 0.00527 0.03872 0.96128 0 0.02434 0.97566 0 0.01837 0.98163 0 0.01077 0.96534 0.02389 0.01976 0.98024 0 0.02478 0.97522 0 0.01289 0.98711 0 0.01124 0.98876 0 0.00961 0.92654 0.06385 0.01855 0.98145 0 0.04182 0.95818 0 0.03635 0.96365 0 0.03184 0.85299 0.11517 0.00573 0.22454 0.76973 0.00229 0.13408 0.86363 0.00093 0.07687 0.9222 0 0.00209 0.9979 0 0.00221 0.99779 0.00002 0.00829 0.99169 0.00065 0.09561 0.90374 0.01589 0.98411 0 0.00991 0.99009 0 0.00443 0.99557 0 0.01314 0.98686 0 0.44616 0.55384 0 0.68628 0.31372 0 1 0 0 0.98957 0.01043 0 0.98792 0.01208 0 0.97521 0.02479 0 1 0 0 1 0 0 0.43647 0.55668 0.00685 0.28318 0.71682 0 0.14261 0.83132 0.02607 0.13582 0.86418 0 0.1056 0.8944 0 0.03629 0.96371 0 0.00279 0.27949 0.71772 0.00143 0.12575 0.87282 0.00011 0.02912 0.97078 0.00001 0.00592 0.99407 0.00002 0.00703 0.99295 0.00012 0.02351 0.97637 0.00064 0.06857 0.93078 0.00139 0.24954 0.74907 0.00784 0.99216 0 0.01713 0.94333 0.03954 0.057 0.943 0 0.05842 0.91425 0.02733 0.03722 0.96278 0 0.03722 0.96278 0
- 80030 12 11 121 127 0.05559 0.94441 0 0.07314 0.80951 0.11736 0.15147 0.84853 0 0.18374 0.81626 0 0.29586 0.70414 0 1 0 0 0.99416 0.00584 0 0.97643 0.02343 0.00014 1 0 0 1 0 0 0.9949 0.0051 0 1 0 0 0.94805 0.05195 0 1 0 0 0.95133 0.04864 0.00003 0.5887 0.4113 0 0.1536 0.8464 0 0.00204 0.10279 0.89517 0.00008 0.0541 0.94582 0.00026 0.07795 0.92179 0.00438 0.43379 0.56184 0.01809 0.98191 0 0.02748 0.97252 0 0.01871 0.98129 0 0.01907 0.98093 0 0.02263 0.97737 0 0.00829 0.442 0.54971 0.00066 0.13393 0.86541 0.00216 0.13426 0.86358 0.00138 0.08847 0.91015 0.0027 0.12535 0.87195 0.0035 0.21603 0.78047 0.02032 0.49739 0.48228 0.05 0.95 0 0.06282 0.92949 0.00769 0.06502 0.92616 0.00882 0.0801 0.9199 0 0.08891 0.91109 0 0.08646 0.91354 0 0.0813 0.9187 0 0.08568 0.91432 0 0.2608 0.7392 0 0.29967 0.70033 0 0.36423 0.63577 0 0.45359 0.53993 0.00649 0.48542 0.51458 0 1 0 0 1 0 0 0.48916 0.50519 0.00566 0.38395 0.61605 0 0.08177 0.91823 0 0.06985 0.90434 0.02581 0.01758 0.98242 0 0.00242 0.99758 0 0.00914 0.99086 0 0.04127 0.95873 0 0.05606 0.93267 0.01127 0.06201 0.93799 0 0.06201 0.93799 0

fam id1 id2 piq1 piq2 ibd0m1 ibd1m1 ibd2m1 ibd0m2 ibd1m2 ibd2m2 ….

*phenotypes* *ibd probabilities to calculate pihats at different locations*

# DZ with pi-hat -> FEQ



# Definition Variables

- Represented by diamond in diagram
- Changes likelihood for every individual in the sample according to their value for that variable

```
#define nvar 1
#NGroups 1

DZ / SIBS genotyped
 Data NInput=182 Maxrec=1500
 Rectangular File=piqibd.rec
  Labels fam id1 id2 piq1 piq2
 ibd0m1 ibd1m1 ibd2m1 ibd0m2 ibd1m2 ibd2m2 ....
 ibd0m59 ibd1m59 ibd2m59
  Select  piq1 piq2  ibd0m1  ibd1m1  ibd2m1 ;
  Definition    ibd0m1  ibd1m1  ibd2m1 ;
 Begin Matrices;
  X Lower nvar nvar free ! residual familial F
  Z Lower nvar nvar free ! unshared environment E
  L Full nvar 1 free     ! qtl effect Q
  G Full 1 nvar free     ! grand means
  H Full 1 1             ! scalar, .5
  K Full 3 1             ! IBD probabilities (Merlin)
  J Full 1 3             ! Coefficients 0,.5,1 for pihat
 End Matrices;
```

FEQmodel_Pihat1_DZibd.mx

```
   Specify K    ibd0m1  ibd1m1  ibd2m1 ;
    Matrix H .5
   Matrix J 0 .5 1
    Start ..
  Begin Algebra;
    F= X*X';                    ! residual familial var
    E= Z*Z';                    ! unique environmental var
    Q= L*L';                    ! variance due to QTL
    V= F+Q+E;                   ! total variance
    T= F|Q|E;                   ! parameters in 1 matrix
    S= F%V| Q%V| E%V;           ! standardized var components
   P= J*K;                      ! estimate of pi-hat
  End Algebra;
 Means  G| G ;
 Covariance       F+Q+E | F+P@Q
                  F+P@Q | F+Q+E ;
  Option Multiple Issat
 End

  Drop L 1 1 1              !test significance of QTL effect
 Exit
```

FEQmodel_Pihat1_DZibd.mx

# Practical Pi-hat

- Mx script: FEQmodel_Pihat1_DZibd.mx

- Choose a position, run model
- Fit submodel
- Add -2LnLL to Excel spreadsheet

faculty\hmaes\a20\maes\MxLinkage\

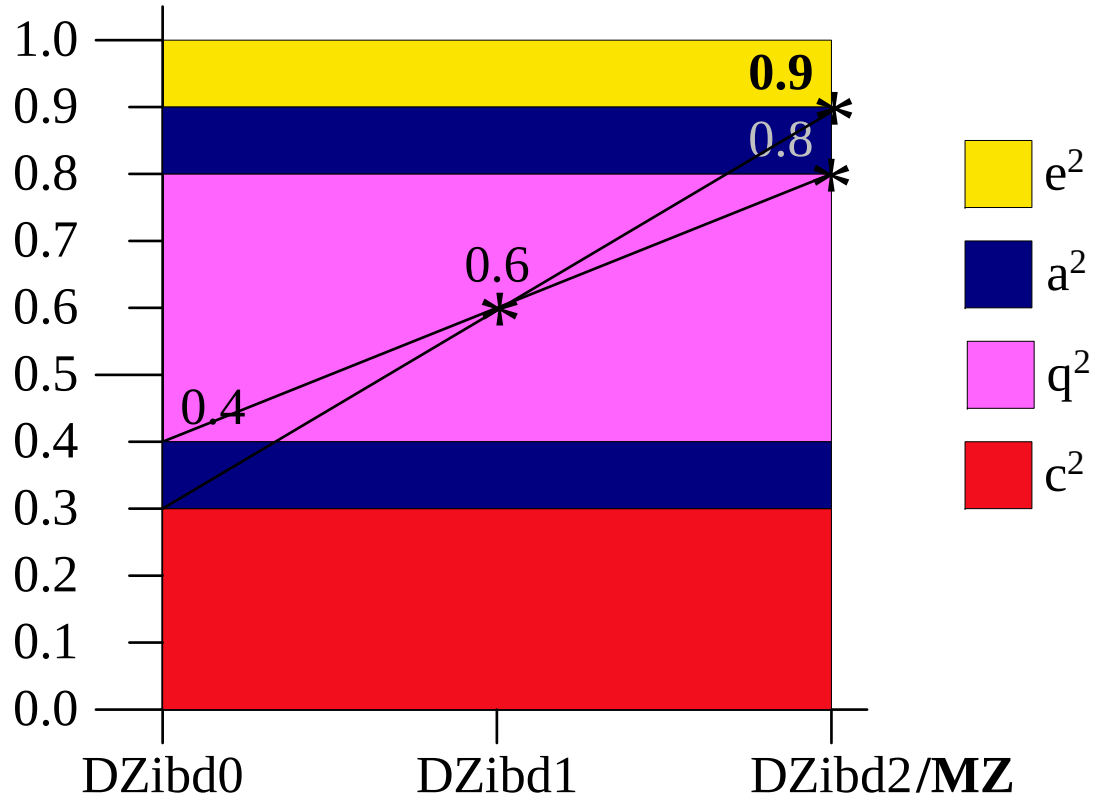
# Test for linkage

- Drop Q from the model
- Note
  - although you will have to run your linkage analysis model many times (for each marker), the fit of the sub-model (or base-model) will always remain the same
  - So… run it once and use the command **`Option Sub=<-2LL>,<df>`**

# Using MZ twins in linkage

- MZ pairs will not contribute to your linkage signal
  - BUT correctly including MZ twins in your model allows you to partition F in A and C or in A and D
  - AND if the MZ pair has a (non-MZ) sibling the 'MZ-trio' contributes more information than a regular (DZ) sibling pair – but less than a 'DZ-trio'
  - MZ pairs that are incorrectly modeled lead to spurious results

# DZ ibd0,1,2 & MZ correlations



# Running a loop (Mx Manual page 52)

- Include a loop function in your Mx script
  - Analyze all markers consecutively
- At the top of the loop
  - `#loop $<number> start stop increment`
    - #loop $nr 1 59 1
- Within the loop
  - One file per chromosome, multiple markers
    - `Select piq1 piq2 ibd0m$nr ibd1m$nr ibd2m$nr`
  - One file per marker, multiple files
    - `Rectangular File =piq$nr.rec`
- At the end of the loop
  - `#end loop`

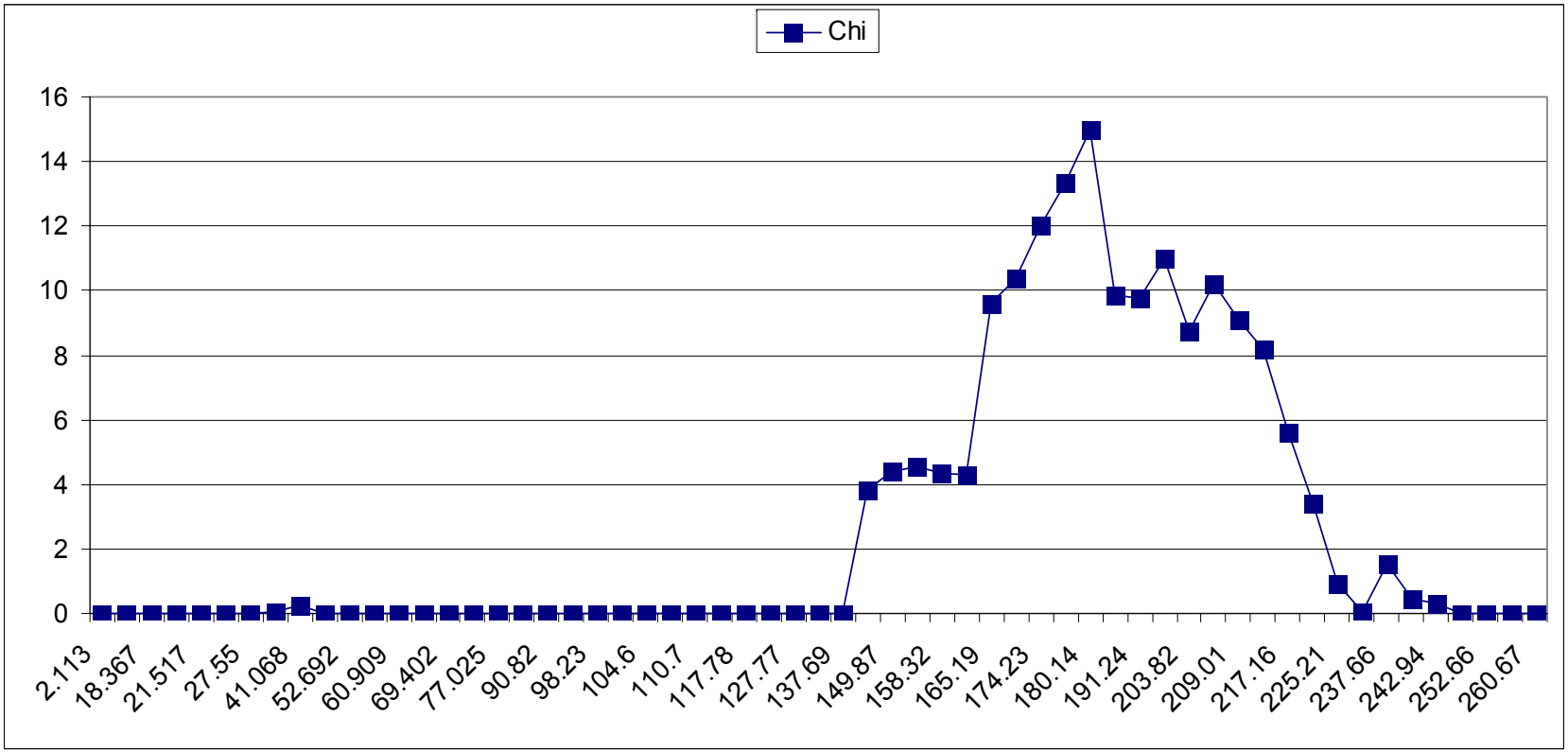
```
#loop $nr 1 59 1
#define nvar 1
#NGroups 1

DZ / SIBS genotyped
 Data NInput=182 Maxrec=1500
 Rectangular File=piqibd.rec
  Labels fam id1 id2 piq1 piq2 ....
  Select  piq1 piq2  ibd0m$nr  ibd1m$nr  ibd2m$nr ;
  Definition    ibd0m$nr  ibd1m$nr  ibd2m$nr ;
 Begin Matrices;
  X Lower nvar nvar free ! residual familial F
  Z Lower nvar nvar free ! unshared environment E
  L Full nvar 1 free     ! qtl effect Q
  G Full 1 nvar free     ! grand means
  H Full 1 1             ! scalar, .5
  K Full 3 1             ! IBD probabilities (Merlin)
  J Full 1 3             ! Coefficients 0,.5,1 for pihat
 End Matrices;
```

FEQmodel_Pihat1-59_DZibd.mx

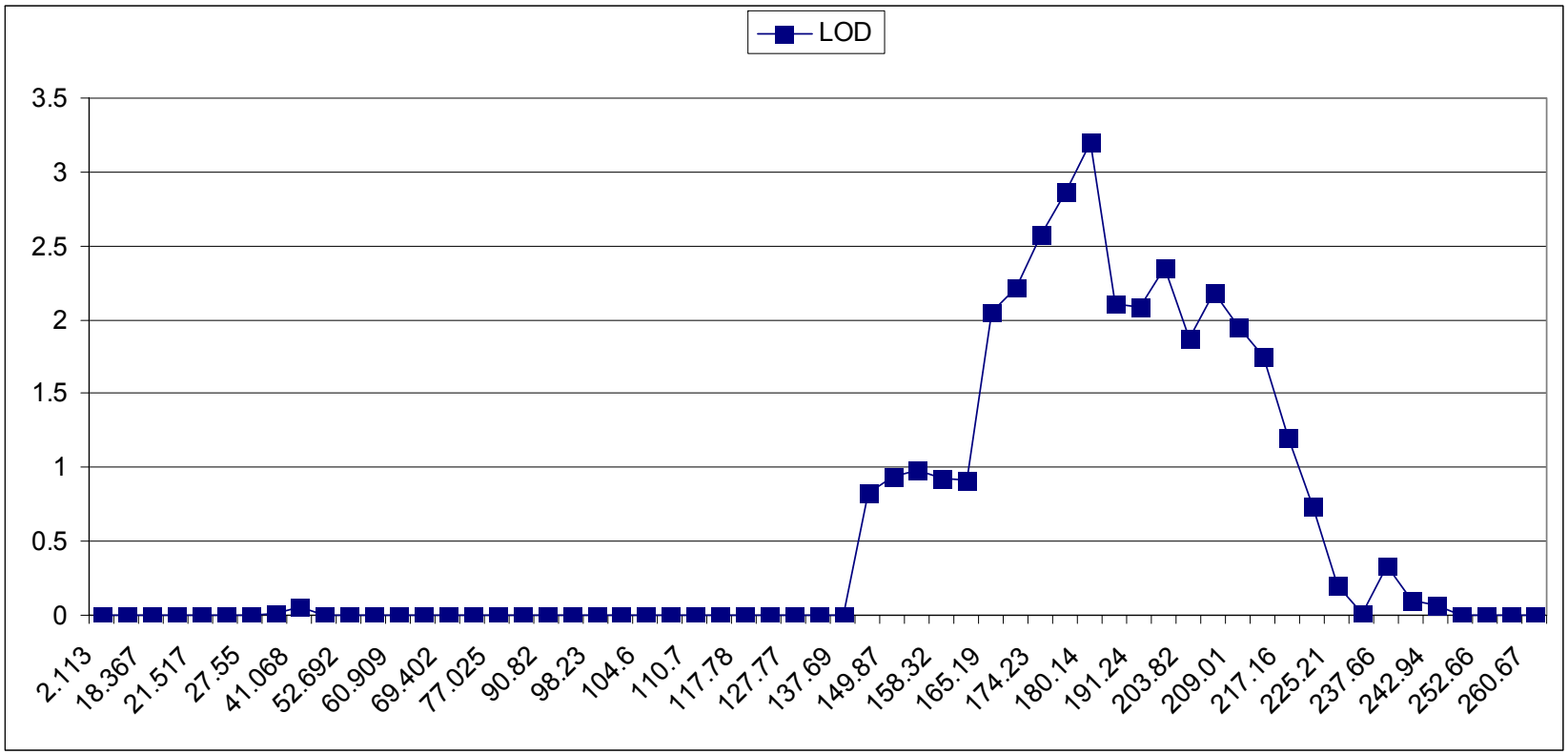
```
Specify K    ibd0m$nr  ibd1m$nr  ibd2m$nr ;
 Matrix H .5
 Matrix J 0 .5 1
 Start ..
Begin Algebra;
 F= X*X';                    ! residual familial var
 E= Z*Z';                    ! unique environmental var
 Q= L*L';                    ! variance due to QTL
 V= F+Q+E;                   ! total variance
 T= F|Q|E;                   ! parameters in 1 matrix
 S= F%V| Q%V| E%V;           ! standardized var components
 P= J*K;                     ! estimate of pi-hat
End Algebra;
Means  G| G ;
Covariance      F+Q+E | F+P@Q_
                F+P@Q | F+Q+E ;
 Options ..
 Option Sub=7203.35,853     ! likelihood, df from FE model
Exit
#end loop
```

FEQmodel_Pihat1-59_DZibd.mx

# Pi-hat Results



# LOD=(Univariate)$\Delta\chi^2/4.61$



# Model Free Linkage

- No need to specify mode of inheritance
- Models phenotypic and genotypic similarity of relatives
- Expression of phenotypic similarity as a function of IBD status