

[Copy folder (and subfolders)]

- F:\sarah\linkage2



Linkage in Mx

Sarah Medland

[In this session:]

- Preparing your data for variance components linkage in Mx
- Showing the progression to linkage via path models
- Running linkage in Mx
 - Introducing the loop function
 - Sibship size 5
 - Efficient algebra

[From Merlin to Mx]

Merlin gives us...

```

FAMILY ID1 ID2 MARKER P0 P1 P2
60007 31 31 0.000 0.0 0.0 1.0
60007 41 31 0.000 1.0 0.0 0.0
60007 41 41 0.000 0.0 0.0 1.0
60007 1 31 0.000 0.0 1.0 0.0
60007 1 41 0.000 0.0 1.0 0.0
60007 1 1 0.000 0.0 0.0 1.0
60007 2 31 0.000 0.0 1.0 0.0
60007 2 41 0.000 0.0 1.0 0.0
60007 2 1 0.000 0.18899 0.54215 0.26886
60007 2 2 0.000 0.0 0.0 1.0

```

Mx wants something like...

Pihats

0.536	0.492	0.5	0.5	0.039	0.5	0.5	0.5	0.5	0.5	-0.81	-0.98	0.303	.	.
0.530	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.156	-0.26	.	.	.
0.003	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	-0.12	-0.67	.	.	.
0.464	0.039	0.5	0.5	0.505	0.5	0.5	0.5	0.5	0.5	0.585	-0.40	0.144	.	.
0.003	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	-0.42	-0.79	.	.	.
0.039	0.074	0.960	0.5	0.962	0.003	0.5	0.037	0.5	0.5	3.156	-0.51	2.175	0.162	.
0.940	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	2.220	0.867	.	.	.
0.064	0.909	0.5	0.5	0.133	0.5	0.5	0.5	0.5	0.5	-1.33	-1.95	-0.94	.	.
0.041	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.733	1.887	.	.	.
0.536	0.003	0.498	0.5	0.463	0.042	0.5	0.502	0.5	0.5	-0.60	-1.17	2.092	0.854	.

Phenotypes

[From Merlin to Mx]

- Different ways to go about this
 - Shell or Perl scripts in Unix/Linux 👍 😞
 - Alsort (for pairwise data) 👍 😊
 - SPSS, SAS, R etc 👎 😊

[Alsort.exe]

- Takes an all-possible pairs approach rather than a full sib-ship approach
 - If a family has a sibship of 2 then 1 pair
 - If a family has a sibship of 3 then 3 pairs
 - If a family has a sibship of 4 then 6 pairs
- You can run alsort and then convert to a full sib-ship approach

Alsort.exe – to go through at home

Usage: alsort <inputfile> <outfile> [-vfpm] [-c] [-i] [-t] [-x <id1> ...]

-v Verbose (implies -vfpm)

-vf Print family ID list

-vp Print marker positions

-vm Print missing p-values

-c Create output file per chromosome

-i Include 'self' values (id1=id2)

-x Exclude list; id-values separated by spaces

-t Write tab as separator character

Thanks to Danielle Posthuma for this slide and this practical example

Practical Alsort.exe

There is a copy of alsort in the folder:
pract-alsort

Open a dos prompt, go to the directory
where alsort.exe is and type

```
Alsort test.ibd sorted.txt -c -x 31 41 -t
```

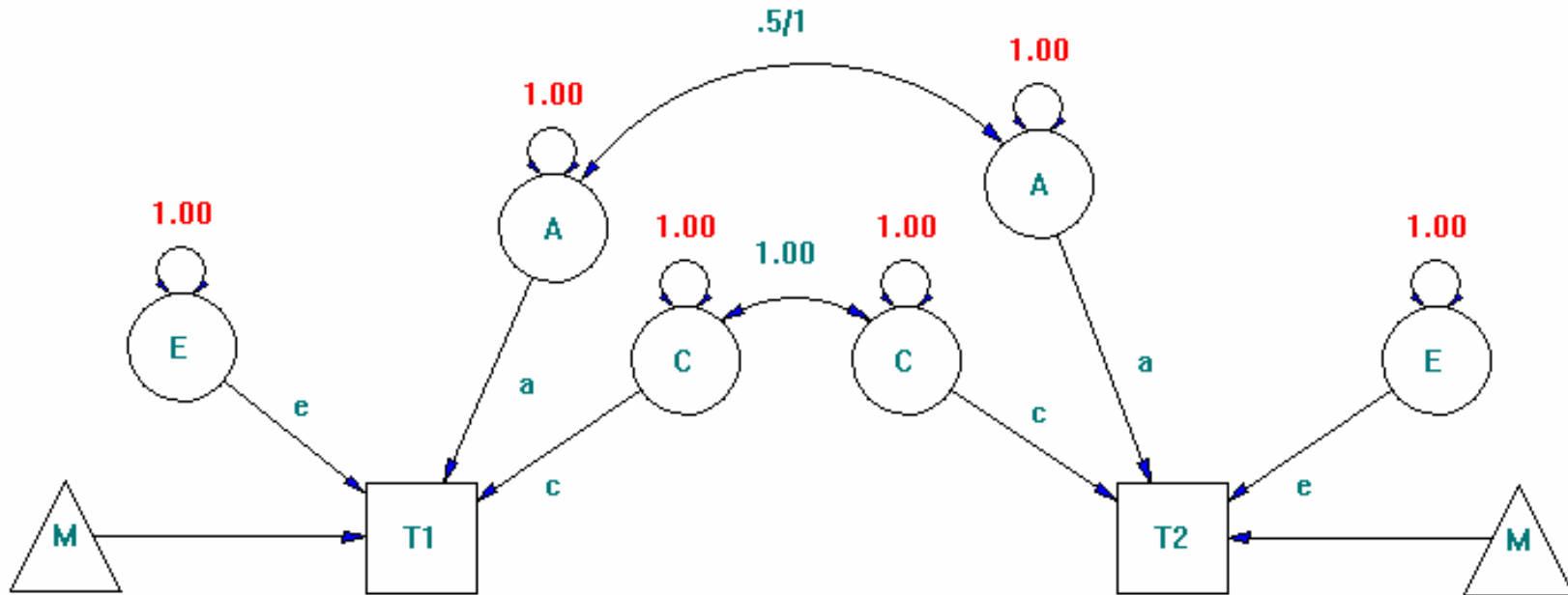
Thanks to Danielle Posthuma for this slide and this practical example

[Once you have your data...]

- Estimates of QTL effects can easily be incorporated into ACE/ADE models
- 'Simple' extension of the path model and script

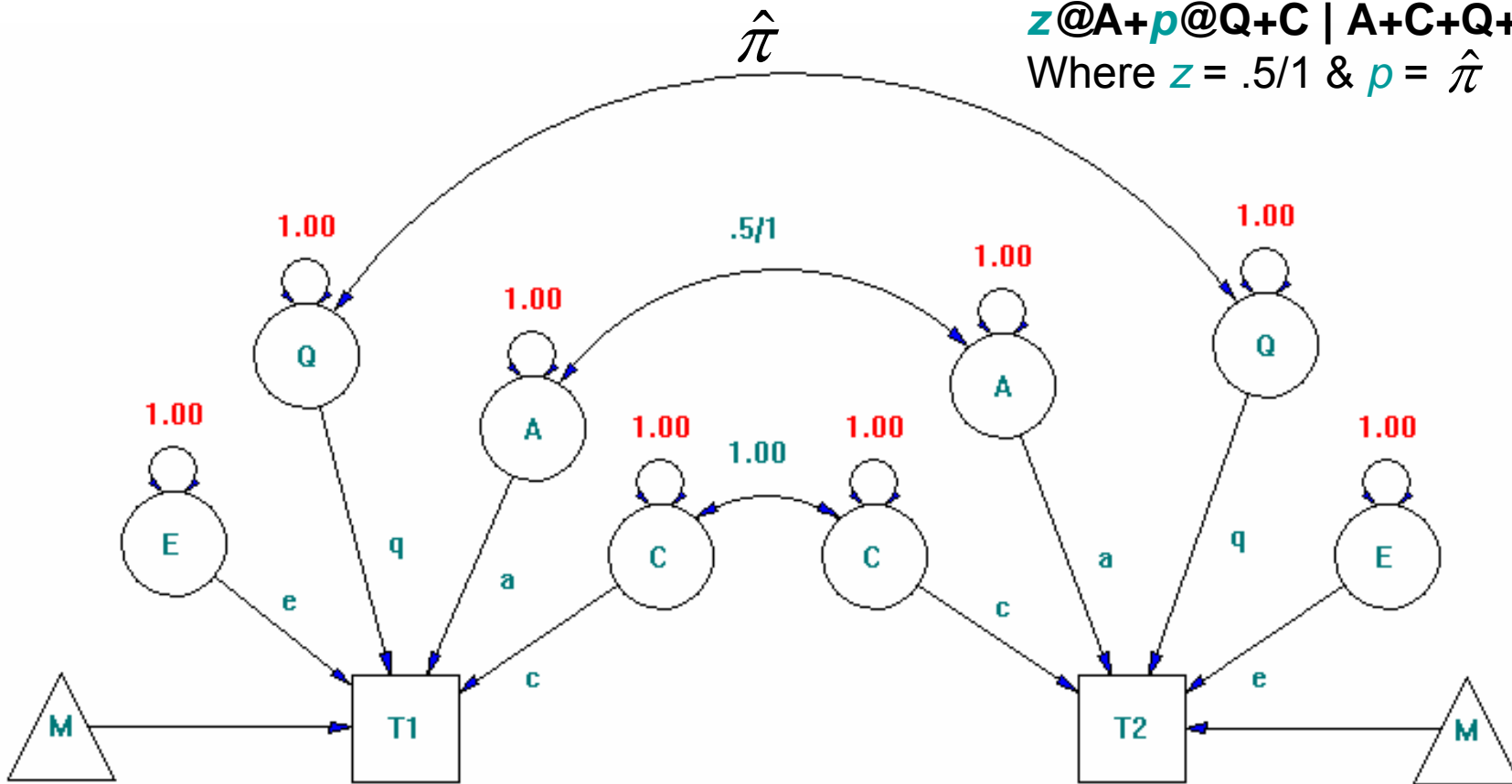
[ACE path model]

$A+C+E \mid z@A+C_$
 $z@A+C \mid A+C+E ;$
 Where $z = .5/1$



[ACQE path model]

$A+C+Q+E | z@A+p@Q+C_$
 $z@A+p@Q+C | A+C+Q+E ;$
 Where $z = .5/1$ & $p = \hat{\pi}$

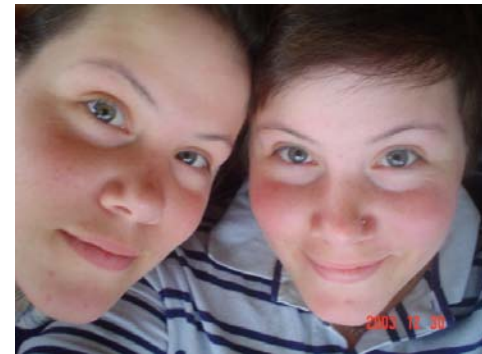


[Test for linkage]

- Drop Q from the model
- Note
 - although you will have to run your linkage analysis model many times 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 that are
incorrectly modelled
lead to spurious
results



Using MZ twins in linkage

- An MZ pair will not contribute to your linkage signal
 - BUT correctly including MZ twins in your model allows you to partition A and C or A and E
 - AND if the MZ pair have a (non-MZ) sibling the 'MZ-trio' contributes more information than a regular (DZ) sibling pair – but less than a 'DZ-trio'

[Running linkage]

- Multiple jobs
 - Run 'at the Marker'
 - Run 'at a Grid'
 - Approx 3500 cM
 - Every 1/2/5/ cM?
- So... we want the analyses to be efficient

[Running a loop]

- Pre-prepare your data files
 - One per-chromosome or one per-marker
- Include a loop function in your mx script to select the data you want to analyse

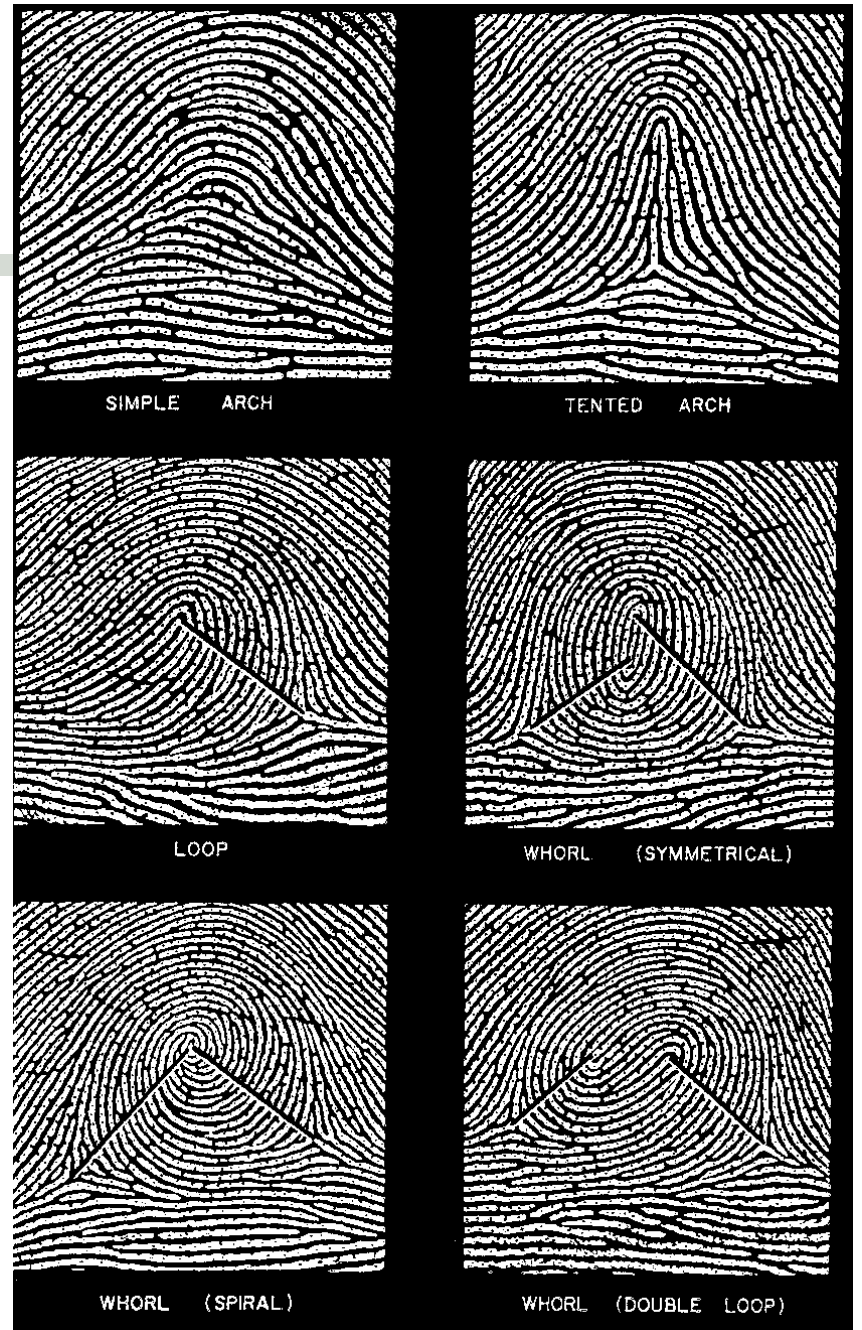
[Running a loop (page 52)]

- At the top of the loop
 - #loop \$<number> start stop increment
 - #loop \$file 39 59 2
- Within the loop
 - Rectangular file =grid\$file`.rec
 - Rectangular file =grid39.rec
- At the end of the loop
 - #endloop

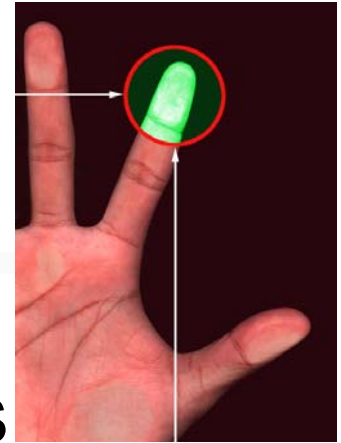
[Practical Example

Absolute Finger Ridge Count

- Sum of total number of ridges between the triradii (deltas) and cores of all ten fingers



Practical Example



- Australian adolescent twins and sibs
- Sibships of up to 5 siblings



Two 10cM genome scans
interleaved = Approx 5cM
scan

Covariance structure

- It is important to correctly structure the $\hat{\pi}$ coefficients for the QTL variance component

	a	b	c	d	e
a	1	$\hat{\pi}_{a,b}$	$\hat{\pi}_{a,c}$	$\hat{\pi}_{a,d}$	$\hat{\pi}_{a,e}$
b	$\hat{\pi}_{a,b}$	1	$\hat{\pi}_{b,c}$	$\hat{\pi}_{b,d}$	$\hat{\pi}_{b,e}$
c	$\hat{\pi}_{a,c}$	$\hat{\pi}_{b,c}$	1	$\hat{\pi}_{c,d}$	$\hat{\pi}_{c,e}$
d	$\hat{\pi}_{a,d}$	$\hat{\pi}_{b,d}$	$\hat{\pi}_{c,d}$	1	$\hat{\pi}_{d,e}$
e	$\hat{\pi}_{a,e}$	$\hat{\pi}_{b,e}$	$\hat{\pi}_{c,e}$	$\hat{\pi}_{d,e}$	1

Covariance structure

- In the Mx script: cov_loop_linkage.mx

```
Labels pab pac pad pae pbc pbd pbe pcd pce pde arc1 arc2 arc3 arc4 arc5  
select pab pac pad pae pbc pbd pbe pcd pce pde arc1 arc2 arc3 arc4 arc5 ;  
definiton pab pac pad pae pbc pbd pbe pcd pce pde ;
```

```
      P Full 1 1          Specify P pab  
      R Full 1 1          Specify R pac  
      S Full 1 1          Specify S pad  
      T Full 1 1          Specify T pae  
      U Full 1 1          Specify U pbc  
      V Full 1 1          Specify V pbd  
      W Full 1 1          Specify W pbe  
      X Full 1 1          Specify X pcd  
      Y Full 1 1          Specify Y pce  
      Z Full 1 1          Specify Z pde
```

Covariance structure

- In the Mx script: cov_loop_linkage.mx

Covariance

```
A+C+E+D+Q | H@A+C+I@D+P@Q | H@A+C+I@D+R@Q | H@A+C+I@D+S@Q | H@A+C+I@D+T@Q _
H@A+C+I@D+P@Q | A+C+E+D+Q | H@A+C+I@D+U@Q | H@A+C+I@D+V@Q | H@A+C+I@D+W@Q _
H@A+C+I@D+R@Q | H@A+C+I@D+U@Q | A+C+E+D+Q | H@A+C+I@D+X@Q | H@A+C+I@D+Y@Q _
H@A+C+I@D+S@Q | H@A+C+I@D+U@Q | H@A+C+I@D+X@Q | A+C+E+D+Q | H@A+C+I@D+Z@Q _
H@A+C+I@D+T@Q | H@A+C+I@D+W@Q | H@A+C+I@D+Y@Q | H@A+C+I@D+Z@Q | A+C+E+D+Q ;
```

[Your task...]

- You have been given data for a section of a chromosome
- Edit the loop in the script `cov_loop_linkage.mx`

Based on where you are sitting

- Run linkage for your assigned grids

[Huh?!?]

Front of the room

Grid
39 – 43
Which
maps to
190-
210cM

Grid
44 – 48
Which
maps to
215-
235cM

Grid
49 – 53
Which
maps to
240-
245cM

Grid
54 – 59
Which
maps to
250-
290cM



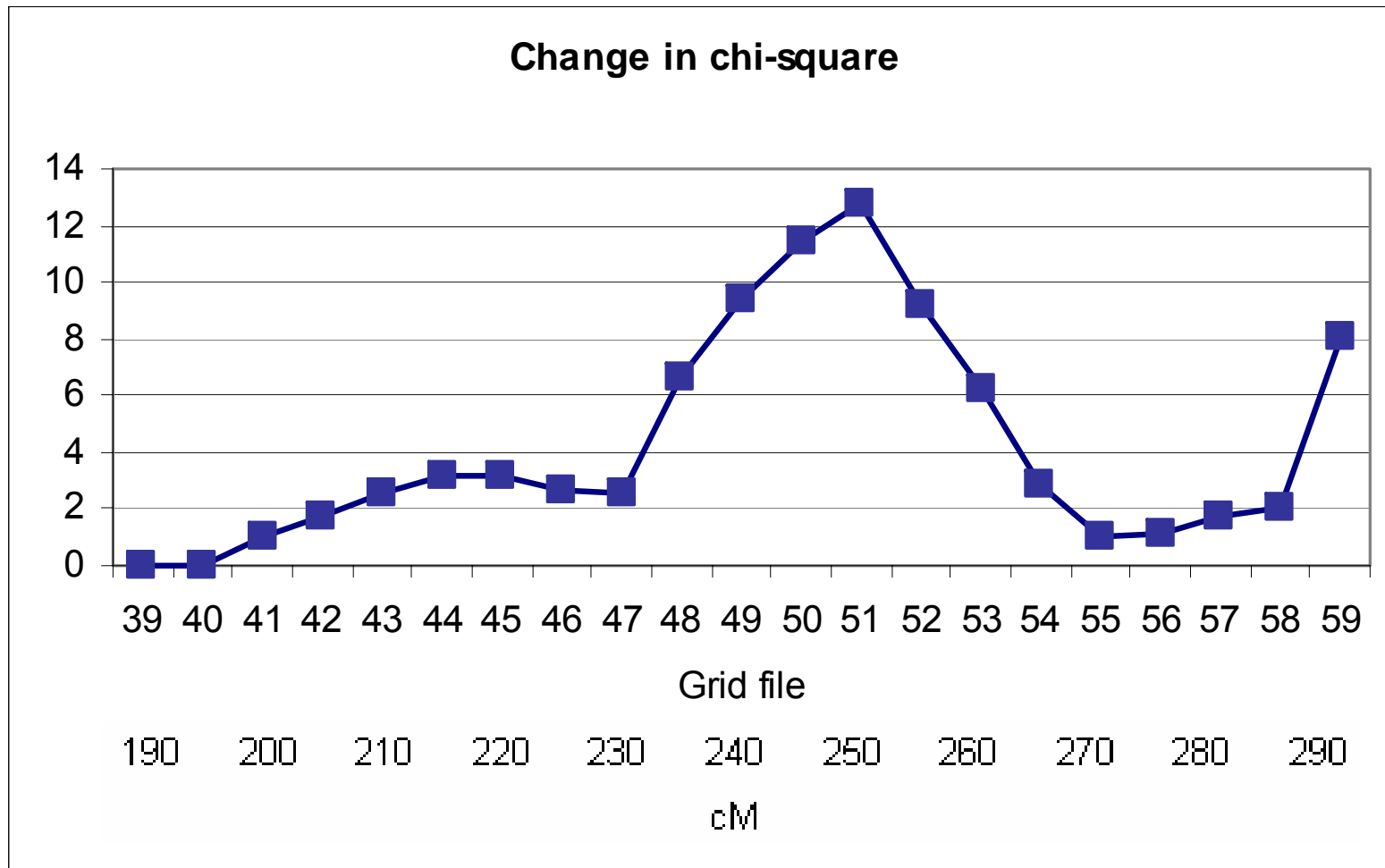
[Your task...]

- Edit the loop in the script `cov_loop_linkage.mx`

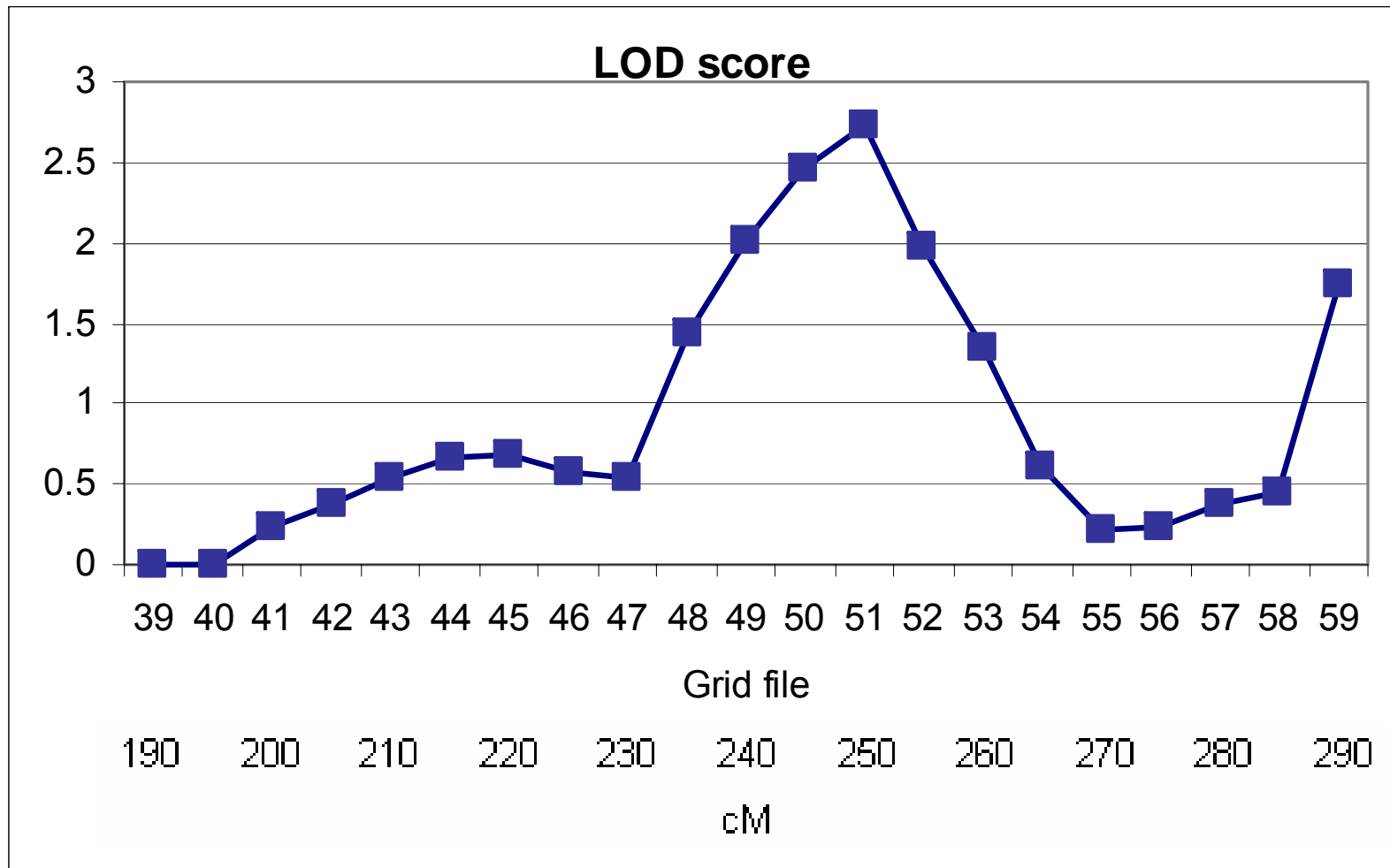
Based on where you are sitting

- Run linkage for your assigned grids
- Note the change in chi-square for each grid
- Come and add your results to the graph on the excel sheet up the front

One I prepared earlier...



$$\text{LOD} = (\text{Univariate}) \Delta\chi^2 / 4.67$$



Significance

- What is the p-value at 250cM (Grid 51)?
 - $\Delta\chi^2 = 12.773$, LOD = 2.74
 - Mixture distribution (MORE on this next year)
 - $\frac{1}{2}$ point mass zero
 - $\frac{1}{2} \chi^2$ df=1
 - Called $\chi_{0.1}^2$
 - From a practical point of view divide the $p\chi_1^2$ by 2
 - So...p=.000176
 - Lander and Kruglyak
 - LOD = 3.6, Chi-sq = 16.7, $P = 0.000022$

[Alternatively]

- Estimate significance taking into account the informative-ness of your sample and your genotypic data
 - Simulate 1,000-10,000 null replicates
 - MORE on this next year

[From a practical point of view]

- Speed is important!!!
- Have a look at `kron_loop_linkage.mx`

Covariance structure

- In the Mx script: kron_loop_linkage.mx

```
Labels pab pac pad pae pbc pbd pbe pcd pce pde arc1 arc2 arc3 arc4 arc5
select pab pac pad pae pbc pbd pbe pcd pce pde arc1 arc2 arc3 arc4 arc5 ;
definition pab pac pad pae pbc pbd pbe pcd pce pde ;
```

```
X stand nsib nsib          ?A coefficients
Y unit nsib nsib          ?C coefficients
W stand nsib nsib          ?D coefficients
L stand nsib nsib fixed    ?Q coefficients - pihats
Z ident nsib nsib fixed    ?E coefficients
End Matrices;
```

```
SP L pab pac pbc pad pbd pcd pae pbe pce pde ;
Value .5 X 2 1 - X 5 4 ;
Value .25 W 2 1 - W 5 4 ;
Means SQM ;
Covariances
(X@A)+(X@C)+(W@D)+(Z@E)+(L@Q) ;
```

[From a practical point of view]

- Speed is important!!!
- Have a look at `kron_loop_linkage.mx`
- Run it – how does it compare
 - In terms of speed?
 - Do you get the same results?



For more on these topics

Come back next year 😊

or in the meantime read the
recommended readings for next
year