Variance Components Models for Heritability and QTL Mapping

Mendel Analysis Option polygenic_qtl

Page 131 of the Manual

(also option kinship_matrices or simwalk)

Input files:

(1) control.in  (2) map.in  (3) pedigree.in  (4) locus.in
(5) VAR.IN
(6) Coefficients.in
(7) kekey.txt
Polygenic analysis: Variance Component Models can be used to determine the heritability of a trait

Simple example: Assume the genes and environmental effects are independent, that the environmental effects are independent across individuals and that the polygenic effect is due to many independent additive acting genes of approximately the same effect size.

Let \( Y_i = \mu + \beta^T a_i + g_i + e_i \) where \( \mu \) is the population mean, \( a \) are the “environmental” predictor variables, \( g \) is the polygenic effect, and \( e \) is the residual error.

The variance of the trait can be partitioned \( V_p = V_g + V_e \) and the covariance is \( \text{cov}(Y_i, Y_j) = 2\Phi_{ij} V_g \) where \( \Phi_{ij} \) is the theoretical kinship coefficient. Heritability equals \( V_g / V_p \).
The degree of correlation between two relatives depends on the theoretical kinship coefficient

- the probability that two alleles, at a randomly chosen locus, one chosen randomly from individual $i$ and one from $j$ are identical by descent.
- The kinship coefficient does not depend on the observed genotype data.
Example of polygenic analysis - finger prints

- Total finger ridge count for the right hand.
- Families selected because proband had abnormally high ridge count.
Mendel Control File

PEDIGREE_FILE = Ped19a.in
LOCUS_FILE = Locus19a.in  ! needed to identify proband
VARIABLE_FILE = Variable19a.in ! describes the variables
OUTPUT_FILE = Mendel_heritability.out

QUANTITATIVE_TRAIT = Right  ! right = \mu + \beta*sex
PREDICTOR = Grand :: Right  ! specifies \mu in the model
PREDICTOR = Sex :: Right  ! sex: -1 if female 1 if male
ANALYSIS_OPTION = Polygenic_qtl

COVARIANCE_CLASS = Additive  ! no dominance in polygenic
COVARIANCE_CLASS = Environmental  ! independent
! no QTL in this model!

PROBAND = P
PROBAND_FACTOR = PROBAND
### SUMMARY FOR MEAN PARAMETERS

<table>
<thead>
<tr>
<th>TRAIT</th>
<th>PARAMETER</th>
<th>PREDICTOR</th>
<th>ESTIMATE</th>
<th>STD ERR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right 1</td>
<td>GRAND (µ)</td>
<td>66.1033</td>
<td>2.4510</td>
<td></td>
</tr>
<tr>
<td>Right 2</td>
<td>FEMALE (-β)</td>
<td>-3.1805</td>
<td>1.5546</td>
<td></td>
</tr>
<tr>
<td>Right 3</td>
<td>MALE (β)</td>
<td>3.1805</td>
<td>1.5546</td>
<td></td>
</tr>
</tbody>
</table>

### SUMMARY FOR ADDITIVE VARIANCE (polygenic)

<table>
<thead>
<tr>
<th>TRAIT</th>
<th>PARAMETER</th>
<th>ESTIMATE</th>
<th>STD ERR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right 4</td>
<td></td>
<td><strong>687.2783</strong></td>
<td>73.3659</td>
</tr>
</tbody>
</table>

### SUMMARY FOR ENVIRONMENTAL VARIANCE

<table>
<thead>
<tr>
<th>TRAIT</th>
<th>PARAMETER</th>
<th>ESTIMATE</th>
<th>STD ERR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right 5</td>
<td></td>
<td><strong>33.3144</strong></td>
<td>10.8574</td>
</tr>
</tbody>
</table>

Heritability = 687.2783/(687.2783+33.3144) = 0.95
QTL mapping using a variance component model

Use a variance component model to test whether the covariance among relatives trait values is correlated with the identity by descent (IBD) sharing at a locus.

We use the conditional kinship coefficient as a measure of IBD: The conditional kinship coefficient is the probability that a gene chosen randomly from person \( i \) at a specific locus matches a gene chosen randomly from person \( j \) given the available genotype information at markers.
Mathematically:

\[ Y_i = \mu + \beta^T a_i + g_i + q_i + e_i \]

where \( \mu \) is the population mean, \( a \) are the “environmental” predictor variables, \( q \) is the major trait locus, \( g \) is the polygenic effect, and \( e \) is the residual error.
The variance of Y is:

$$Var(Y) = V_q + V_g + V_e$$

and for relatives $i$ and $j$:

$$Cov(Y_i, Y_j) = 2\hat{\Phi}_{ij} V_q + 2\Phi_{ij} V_g$$

where

$V_q =$ additive genetic variance due to the QTL, $V_g =$ polygenic variance, $\Phi_{ij} =$ the theoretical kinship coefficient for $i$ and $j$, $\hat{\Phi}_{ij} =$ the conditional kinship coefficient for $i$ and $j$ at a map location

The dominance variance due to the QTL is assumed to be negligible here ($V_d = 0$).

As the map distance increases, the covariance of the trait values becomes less dependent on the value of the conditional kinship coefficient at the map location and so the value of QTL variance component will decrease.
Variance component methods of linkage analysis example overview:

1. Estimate the IBD sharing at specified locations along the genome using marker data.
2. Estimate the variance components $V_g$ and $V_e$, under the null model by maximizing the likelihood.
3. Given the IBD sharing, estimate the variance components, $V_q$, $V_g$ and $V_e$ by maximizing the likelihood using the IBD sharing at specified map positions $Z$.
4. Calculate the location score for each map position $Z$,

$$\log_{10} \left( \frac{L(Z)}{L(Z = \infty)} \right)$$

Identify the map positions where the location score is large.
QTL Example:
The mystery trait example from the Mendel manual:

Besides the usual commands:
ANALYSIS_OPTION = Polygenic_Qtl
VARIABLE_FILE = Variable19b.in
COEFFICIENT_FILE = Coefficient19b.in !ibd info from sibwalk
GRID_INCREMENT = 0.005 !spacing of the map points
! Mean model
QUANTITATIVE_TRAIT = Trait1
PREDICTOR = Grand :: Trait1
PREDICTOR = SEX :: Trait1
PREDICTOR = AGE :: Trait1
PREDICTOR = BMI :: Trait1
! Covariance model
COVARIANCE_CLASS = Additive !polygenic
COVARIANCE_CLASS = Environmental
COVARIANCE_CLASS = Qtl !now specify an additive qtl
Results

- Get a summary file and a full output file
- The summary file looks like:

<table>
<thead>
<tr>
<th>MARKER</th>
<th>MAP</th>
<th>LOCATION</th>
<th>AIC</th>
<th>NUMBER OF FACTORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker01</td>
<td>0.0000</td>
<td>1.5892</td>
<td>6.6816</td>
<td>1</td>
</tr>
<tr>
<td>Marker02</td>
<td>0.0010</td>
<td>1.5679</td>
<td>6.7798</td>
<td>1</td>
</tr>
<tr>
<td>--</td>
<td>0.0050</td>
<td>1.6693</td>
<td>6.3126</td>
<td>1</td>
</tr>
<tr>
<td>--</td>
<td>0.0100</td>
<td>1.8603</td>
<td>5.4329</td>
<td>1</td>
</tr>
<tr>
<td>--</td>
<td>0.0150</td>
<td>2.1112</td>
<td>4.2778</td>
<td>1</td>
</tr>
<tr>
<td>--</td>
<td>0.0200</td>
<td>2.4028</td>
<td>2.9346</td>
<td>1</td>
</tr>
<tr>
<td>Marker03</td>
<td>0.0228</td>
<td>2.5740</td>
<td>2.1463</td>
<td>1</td>
</tr>
<tr>
<td>Marker04</td>
<td>0.0238</td>
<td>2.5757</td>
<td>2.1383</td>
<td>1</td>
</tr>
<tr>
<td>--</td>
<td>0.0250</td>
<td>2.5666</td>
<td>2.1804</td>
<td>1</td>
</tr>
<tr>
<td>--</td>
<td>0.0300</td>
<td>2.4896</td>
<td>2.5351</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ AIC = -2\ln(L(Z)) + 2n \]  The smaller the AIC the better the fit

\[ n = \text{number of parameters} - \text{number of constraints} \]

Factors will be explained in a little while.
The model we have been considering is very simple:

(1) When examining large pedigrees, it may be possible to consider more realistic models for the environmental covariance.

Example: Modeling common environmental effects using a household indicator. \( H = 1 \) if \( i \) and \( j \) are members of the same household and \( H = 0 \) if \( i \) and \( j \) are not members of the same household.

\[
V_P = 2 \hat{\Phi}_{ij} V_q + 2 \Phi_{ij} V_g + H_{ij} V_c + V_e
\]

(2) The variance component model can use multiple quantitative traits as the outcome.
Using more than one quantitative trait in the analysis

- The model extends so that multiple traits can be considered at the same time.
- The phenotypic variance is now a matrix.
- The variance components get more complicated. Instead of one term per variance component, there are \((1+...+n) = (n+1)*n/2\) terms where \(n\) is the number of quantitative traits.
- As an example, consider two traits \(X\) and \(Y\).

\[
\begin{pmatrix}
V_{PX} & V_{PXY} \\
V_{PXY} & V_{PY}
\end{pmatrix}
= \begin{pmatrix}
V_{gX} & V_{gXY} \\
V_{gXY} & V_{gY}
\end{pmatrix} + \begin{pmatrix}
V_{qX} & V_{qXY} \\
V_{qXY} & V_{qY}
\end{pmatrix} + \begin{pmatrix}
V_{eX} & V_{eXY} \\
V_{eXY} & V_{eY}
\end{pmatrix}
\]
For numerical stability, it is better to reparameterize the variances using factor analytic approach

• Factor refers to hidden underlying variables that capture the essence of the data
• Each variance component is parameterized as factors.
• We will illustrate with the additive genetic variance matrix for two traits $X$ and $Y$ (in principle any number of traits or any of the components could have been used).
• There exists a matrix
  \[
  \begin{pmatrix}
  \delta_{q1} & 0 \\
  \delta_{q12} & \delta_{q2}
  \end{pmatrix}
  \]
such that:

\[
V_{qX} = \delta_{q1}^2, V_{qXY} = \delta_{q1} \delta_{q12}, V_{qY} = \delta_{q12}^2 + \delta_{q2}^2
\]
Example of a Reduction in Parameters

Recall the original factor matrix for the QTL

\[
\begin{pmatrix}
\delta_{q1} & 0 \\
\delta_{q12} & \delta_{q2}
\end{pmatrix}
\]

Set \( \delta_{q2} = 0 \)

\[
V_{qX} = \delta_{q1}^2, V_{qXY} = \delta_{q1} \delta_{q12}, V_{qY} = \delta_{q12}^2
\]
Why Factor Analysis?

- Can lead to a model with fewer parameters
- More stable numerically if you gradually build up the number of factors
- Automatic satisfaction of covariance matrix constraints
- Possible evidence for pleiotropy
Modifications to the control file

QUANTITATIVE_TRAIT = Trait1
QUANTITATIVE_TRAIT = Trait2
PREDICTOR = Grand :: Trait1
PREDICTOR = SEX :: Trait1
PREDICTOR = AGE :: Trait1
PREDICTOR = BMI :: Trait1
PREDICTOR = Grand :: Trait2
PREDICTOR = SEX :: Trait2
PREDICTOR = AGE :: Trait2
PREDICTOR = BMI :: Trait2
COVARIANCE_CLASS = Additive
COVARIANCE_CLASS = Environmental
COVARIANCE_CLASS = Qtl
One factor explains the results as well as two

<table>
<thead>
<tr>
<th>MARKER</th>
<th>MAP DIST</th>
<th>LOC</th>
<th>SCORE</th>
<th>AIC</th>
<th>#FACTORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker02</td>
<td>0.0010</td>
<td>1.6492</td>
<td>24.4052</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.0050</td>
<td>1.7558</td>
<td>23.9143</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.0100</td>
<td>1.9508</td>
<td>23.0161</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.0150</td>
<td>2.1981</td>
<td>21.8774</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.0200</td>
<td>2.4761</td>
<td>20.5972</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Marker03</td>
<td>0.0228</td>
<td>2.6344</td>
<td>19.8680</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Marker04</td>
<td>0.0238</td>
<td>2.6370</td>
<td>19.8562</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

...  

<table>
<thead>
<tr>
<th>MARKER</th>
<th>MAP DIST</th>
<th>LOC</th>
<th>SCORE</th>
<th>AIC</th>
<th>#FACTORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker02</td>
<td>0.0010</td>
<td>1.6520</td>
<td>26.3925</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.0050</td>
<td>1.7568</td>
<td>25.9098</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.0100</td>
<td>1.9508</td>
<td>25.0162</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.0150</td>
<td>2.1981</td>
<td>23.8774</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>--</td>
<td>0.0200</td>
<td>2.4761</td>
<td>22.5972</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Marker03</td>
<td>0.0228</td>
<td>2.6344</td>
<td>21.8680</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Marker04</td>
<td>0.0238</td>
<td>2.6370</td>
<td>21.8563</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Summary

Mendel can carry out both classical heritability studies and QTL mapping with multivariate traits.

When with two or more traits we can have models with reduced numbers of factors. Could a single factor explain QTL variance components and be considered evidence of pleiotropic effects? A single factor is consistent with pleiotropy although there may be other explanations a single factor.