Biostat 551: Homework 5

Brenda Price

Due December 4, 2014

1. *[50 points] In homework 4, the Pedstats software was used to analyze the data files 'BMI BIOST551.ped' and 'BMI BIOST551.dat' for heritability estimation of BMI in European pedigrees for different classes of relative pairs. You will now estimate heritability of BMI considering all pedigree individuals together in a joint analysis using mixed linear models with the QTDT software. The two data files for the European pedigrees previously used for the homework 4 BMI heritability analysis with Pedstats (available on the course website) are also in the required format for the QTDT software.*

   (a) *Provide variance component estimates for BMI using a mixed linear model (MLM) with the following two random effects: additive polygenic and unique (or non-shared) environmental effects. Provide an estimate of the narrow sense heritability of BMI from the variance component estimates from this mixed linear model.*

   From the 'regress.tbl' file:
   ```
   Family #1 var-covar matrix terms [2]...[[Ve]][[Vg]]
   Family #1 regression matrix...
          [linear] =
          [2 x 1]      Mu
              2.5   1.000
              2.6   1.000

   Some useful information...
                 df : 26412
     log(likelihood) : 76322.95
          variances :  11.895    7.806
              means :   24.844
   ```

   Additive polygenic variance: $= \sigma_g^2 = 7.806$
   Unique enviromental variance: $= \sigma_e^2 = 11.895$
   $h^2 = \frac{\sigma_g^2}{\sigma_g^2+\sigma_e^2} = \frac{7.806}{7.806+11.895} = 0.396.$

   (b) *Now provide variance components estimates for BMI assuming the following three random effects: additive polygenic, shared household, and unique environmental effects. Provide an estimate for narrow sense heritability of BMI with the variance component estimates from this mixed linear model.*
   From the 'regress.tbl' file:

```
Family #1 var-covar matrix terms [3]...[[Ve]][[Vc]][[Vg]]
Family #1 regression matrix...
       [linear] =
       [2 x 1]      Mu
            2.5   1.000
            2.6   1.000


Some useful information...
            df : 26411
  log(likelihood) : 76240.47
        variances :  14.378   3.304   2.220
            means :  24.841
```

Additive polygenic variance: $= \sigma_g^2 = 2.220$

Shared household variance: $= \sigma_c^2 = 3.304$

Unique enviromental variance: $= \sigma_e^2 = 14.378$

$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = \frac{2.220}{2.220 + 3.304 + 14.378} = 0.112$.

(c) *Is the shared household effect for BMI significant for the mixed linear model in 1(b) above? Provide evidence to support your answer.*

From the output:

```
he following models will be evaluated...
  NULL MODEL
     Means = Mu
 Variances = Ve + Vg


  FULL MODEL
     Means = Mu
 Variances = Ve + Vc + Vg

Testing trait:                              BMI
=============================================

 Allele   df(0) -LnLk(0)   df(V) -LnLk(V)   ChiSq       p

   N/A   26412 76322.95   26411 76240.47   164.96   9e-38   (26415 probands)


Run completed on Fri Nov 28 10:17:48 2014
1 tests carried out

The most significant result refers to:
    Trait: BMI
    ChiSq: 164.956
  p-value: 9e-38
```

The likelihood ratio test comparing the model with additive and environmental effects to the model with additive, household, and environmental effects was significant with a p-value of $9x10^{-38}$ and a chisquare statistic of 165. Therefore, the household effect is significant (we can conclude that there is sufficient evidence to

reject the hypothesis that there is no household effect)

(d) *Now compare your BMI heritability estimates from the mixed linear models with variance components in 1(a) and 1(b) above to the BMI estimates obtained in homework 4 based on different relative pair types. Discuss and provide plausible explanations for any differences and/or similarities that you find in the heritability estimates.*

In homework 4, the heritibility estimates ranged from 0.38 to 0.58 across different relative relationships (with sibling relationships being the highest). The estimate in part (a) (0.396) is in that range which makes sense since both the estimate in part (a) and those from homework 4 assumed no dominance and no household effect. The estimate in part (b) (0.112) accounted for a non-zero household effect, and was found to be lower since σ_c^2 was found to be significantly different from zero, as mentioned in part (c).

2. *[50 points] A quantitative trait Y is influenced by a single autosomal locus, family effects, and unique environmental effects in a population. The autosomal locus that influences the trait is bi-allelic with alleles A and a, and the frequency of allele A is 0.15 . Assume that the population is in Hardy-Weinberg equilibrium (HWE) at the locus. For an individual j from family i in the population, the trait has the following model:*

$$Y_{ij} = 1.5 + 0.4G_{ij} + c_i + \epsilon_{ij}$$

where G_{ij} is the number of copies of the A allele that individual j from family i has, c_i is the family effect on the trait for family i, and ϵ_{ij} is the unique environmental effect for individual j in family i. Assume that G_{ij}, c_i, and ϵ_{ij} are independent with $c_i \sim N(0, \sigma_c^2 = 0.02)$, and $\epsilon \sim N(0, \sigma_e^2 = 0.09)$

(a) *Calculate the expected value and variance of the phenotype Y in the population.*

$$
\begin{aligned}
E[Y] &= E[1.5 + 0.4G + c + \epsilon] \\
&= 1.5 + 0.4E[G] + E[c] + E[\epsilon] \\
&= 1.5 + 0.4E[G] + 0 + 0 \\
&= 1.5 + 0.4\mu_G \\
&= 1.5 + 0.4\left(0.15^2(2) + 2(0.15)(0.85)(1) + 0.85^2(0)\right) \\
&= 1.5 + 0.4 * 0.3 \\
&= 1.62
\end{aligned}
$$

$$
\begin{aligned}
Var\,[Y] &= Var\,[1.5 + 0.4G + c + \epsilon] \\
&= 0 + 0.4^2 Var\,[G] + Var\,[c] + Var\,[\epsilon] \\
&= 0.4^2 Var\,[G] + 0.02 + 0.09 \\
&= 0.4^2 \left(E\left[G^2\right] - \mu_G^2 \right) + 0.02 + 0.09 \\
&= 0.4^2 \left(\left(0.15^2(2^2) + 2(0.15)(0.85)(1^2) + 0.85^2(0^2) \right) - \mu_G^2 \right) + 0.02 + 0.09 \\
&= 0.4^2 \left(0.345 - 0.3^2 \right) + 0.02 + 0.09 \\
&= 0.4^2 \left(0.255 \right) + 0.02 + 0.09 \\
&= 0.151
\end{aligned}
$$

(b) *What is the broad sense heritability (H^2) of the phenotype in the population? What is the narrow sense heritability (h^2) of this phenotype in the population?*

$\text{Var}[G] = 0.255$, $\text{Var}[A] = 0.041$, $\text{Var}[C] = 0.02$, $\text{Var}[\epsilon] = 0.09$, $\text{Var}[D] = 0$.

$$
H^2 = \frac{\beta_1^2 \sigma_G^2}{\sigma_Y^2} = \frac{0.041}{0.151} = 0.271
$$

$$
h^2 = \frac{\sigma_A^2}{\sigma_Y^2} = 0.271
$$

(c) *Consider two individuals j and k from family i. Calculate $Cov(Y_{ij}, Y_{ik})$ (i.e., the covariance of Y_{ij} and Y_{ik}) for the following relationship types for j and k:*

- *j and k are mono-zygotic (MZ) twins reared in the same household.*
- *j and k are full-siblings reared in the same household.*
- *j and k are half-siblings reared in the same household.*
- *j and k are half-siblings reared in different households.*
- *j and k are first-cousins reared in different households.*
- *j and k are unrelated but reared in the same household (e.g., one of the individuals is adopted).*

First,

$$
\begin{aligned}
\sigma_D^2 &= 0 \\
\sigma_e^2 &= 0.09 \\
\sigma_A^2 &= 0.041 \\
\sigma_C^2 &= 0.02
\end{aligned}
$$

- j and k are mono-zygotic (MZ) twins reared in the same household.

$$Cov(C) = \sigma_C^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 0.02 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$Cov(A) = \sigma_A^2 \begin{bmatrix} 1 & 2\theta_{jk} \\ 2\theta_{jk} & 1 \end{bmatrix} = \sigma_A^2 \begin{bmatrix} 1 & 2(1/2) \\ 2(1/2) & 1 \end{bmatrix}$$

$$Cov(\epsilon) = \sigma_e^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0.09 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Cov(D) = \sigma_D^2 \begin{bmatrix} 1 & \Delta_7^{jk} \\ \Delta_7^{jk} & 1 \end{bmatrix} = \sigma_D^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\Omega = 2\Theta\sigma_A^2 + \sigma_D^2\Delta_7 + \sigma_c^2\Phi_C + \sigma_E^2 I$$

$$= \sigma_A^2 \begin{bmatrix} 1 & 2(1/2) \\ 2(1/2) & 1 \end{bmatrix} + \sigma_D^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + 0.02 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + 0.09 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Cov(Y_{ij}, Y_{ik}) = \sigma_A^2 + \sigma_D^2 + \sigma_C^2$$
$$= 0.0408 + 0 + 0.02$$
$$= 0.0608$$

- j and k are full-siblings reared in the same household.

$$\Omega^{jk} = 2\Theta^{jk}\sigma_A^2 + \sigma_D^2\Delta_7^{jk} + \sigma_c^2\Phi_C^{jk} + \sigma_E^2 I$$
$$\omega^{jk} = 2\theta^{jk}\sigma_A^2 + \sigma_D^2\Delta_7^{jk} + \sigma_c^2\Phi_C^{jk} + 0$$
$$Cov(Y_{ij}, Y_{ik}) = 2\left(\frac{1}{4}\right)\sigma_A^2 + \sigma_D^2\left(\frac{1}{4}\right) + \sigma_c^2$$
$$= \frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 + \sigma_c^2$$
$$= 0.0204 + 0 + 0.02$$
$$= 0.0404$$

- j and k are half-siblings reared in the same household.

$$\Omega^{jk} = 2\Theta^{jk}\sigma_A^2 + \sigma_D^2\Delta_7^{jk} + \sigma_c^2\Phi_C^{jk} + \sigma_E^2 I$$
$$\omega^{jk} = 2\theta^{jk}\sigma_A^2 + 0 + \sigma_c^2\Phi_C^{jk} + 0$$
$$Cov(Y_{ij}, Y_{ik}) = 2\left(\frac{1}{8}\right)\sigma_A^2 + \sigma_c^2$$
$$= \frac{1}{4}\sigma_A^2 + \sigma_c^2$$
$$= 0.0102 + 0.02$$
$$= 0.0302$$

- j and k are half-siblings reared in different households.

$$\Omega^{jk} = 2\Theta^{jk}\sigma_A^2 + \sigma_D^2\Delta_7^{jk} + \sigma_c^2\Phi_C^{jk} + \sigma_E^2 I$$
$$\omega^{jk} = 2\theta^{jk}\sigma_A^2 + 0 + 0 + 0$$
$$Cov(Y_{ij}, Y_{ik}) = 2\left(\frac{1}{8}\right)\sigma_A^2$$
$$= \frac{1}{4}\sigma_A^2$$
$$= 0.0102$$

- j and k are first-cousins reared in different households.

$$\Omega^{jk} = 2\Theta^{jk}\sigma_A^2 + \sigma_D^2\Delta_7^{jk} + \sigma_c^2\Phi_C^{jk} + \sigma_E^2 I$$
$$\omega^{jk} = 2\theta^{jk}\sigma_A^2 + 0 + 0 + 0$$
$$Cov(Y_{ij}, Y_{ik}) = 2\left(\frac{1}{16}\right)\sigma_A^2$$
$$= \frac{1}{8}\sigma_A^2$$
$$= 0.0051$$

- j and k are unrelated but reared in the same household (e.g., one of the individuals is adopted).

$$\Omega^{jk} = 2\Theta^{jk}\sigma_A^2 + \sigma_D^2\Delta_7^{jk} + \sigma_c^2\Phi_C^{jk} + \sigma_E^2 I$$
$$\omega^{jk} = 0 + 0 + \sigma_c^2\Phi_C^{jk} + 0$$
$$Cov(Y_{ij}, Y_{ik}) = \sigma_c^2$$
$$= 0.02$$

(d) *Now consider two individuals j and k who are from two different (unrelated) families i and l, respectively, where individual j is from family i and individual k is from family l. Calculate $Cov(Y_{ij}, Y_{lk})$.*

$$\Omega^{jk} = 2\Theta^{jk}\sigma_A^2 + \sigma_D^2\Delta_7^{jk} + \sigma_c^2\Phi_C^{jk} + \sigma_E^2 I$$
$$\omega^{jk} = 0 + 0 + 0 + 0$$
$$Cov(Y_{ij}, Y_{ik}) = 0$$

Appendix: Qtdt output and R code

```
### 1 a-c LR Test Output ###
testarossa:qtdt-2.6.1 brendaprice$ /Users/brendaprice/Documents/QTDT/qtdt-2.6.1/qtdt
-d BMI_BIOST551.dat -p BMI_BIOST551.ped -a- -weg -vegc
QTDT - Quantitative TDT 2.6.1
(c) 1998-2007 Goncalo Abecasis (goncalo@umich.edu)

This program implements tests described by
Abecasis et al, Am J Hum Genet 66:279-292 (2000)
Abecasis et al, Eur J Hum Genet 8:545-551 (2000)
and others

The following parameters are in effect:
          QTDT Data File :     BMI_BIOST551.dat (-dname)
      QTDT Pedigree File :     BMI_BIOST551.ped (-pname)
    QTDT IBD Status File :             qtdt.ibd (-iname)
      Missing Value Code :              -99.999 (-xname)
              Covariates :       USER SPECIFIED (-c{p|s|u|-})
       Association Model :                 NONE (-a[a|d|f|m|o|p|r|t|w|-])
     Full Model Variances :           NON SHARED (-v{e|c|g|n|t|a|d|-})
                         &   COMMON ENVIRONMENT
                         &           POLYGENIC
     Null Model Variances :           NON SHARED (-w{e|c|g|n|t|a|d|-})
                         &           POLYGENIC
  Parent of Origin Effects :               NONE (-o[f|t|m|p|-])
 Monte-Carlo Permutations :                  0 (-m9999)
             Random Seed :              123456 (-r9999)
        Numeric Minimizer :      NELDER AND MEAD (-n[f|n|p])
     Transmission Scoring :       FULL PEDIGREE (-t[n|p])

Additional Options
 --dominance, --snp, --multi-allelic, --deviates, --references,
 --exclude-founder-phenotypes, --p-values, --no-regress-tbl

Online documentation http://www.sph.umich.edu/csg/abecasis/QTDT
Comments, bugs: goncalo@umich.edu

The following models will be evaluated...
  NULL MODEL
      Means = Mu
 Variances = Ve + Vg

  FULL MODEL
      Means = Mu
 Variances = Ve + Vc + Vg

Testing trait:                        BMI
================================================

 Allele   df(0)  -LnLk(0)   df(V) -LnLk(V)   ChiSq         p

    N/A   26412 76322.95   26411 76240.47   164.96    9e-38   (26415 probands)


Run completed on Fri Nov 28 10:17:48 2014
1 tests carried out

The most significant result refers to:
    Trait: BMI
    ChiSq: 164.956
  p-value: 9e-38

Trait: BMI
================================================================
Total Probands: 26415
Family #1 Phenotypes - scores :  23.841  24.936
```

```
NULL HYPOTHESIS
---------------

Family #1 var-covar matrix terms [2]...[[Ve]][[Vg]]
Family #1 regression matrix...
       [linear] =
        [2 x 1]       Mu
            2.5   1.000
            2.6   1.000

Some useful information...
              df : 26412
  log(likelihood) : 76322.95
        variances :  11.895   7.806
            means :  24.844

FULL HYPOTHESIS
---------------

Family #1 var-covar matrix terms [3]...[[Ve]][[Vc]][[Vg]]
Family #1 regression matrix...
       [linear] =
        [2 x 1]       Mu
            2.5   1.000
            2.6   1.000

Some useful information...
              df : 26411
  log(likelihood) : 76240.47
        variances :  14.378   3.304   2.220
            means :  24.841
```

```
####### R code #2
sigc <- 0.02
sige <- 0.09
siga <- 2*(0.15)*(0.85)*0.4^2

muG = (0.15^2*(2) + 2*(0.15)*(0.85)*(1) + 0.85^2*(0))
Ey = 1.5 +0.4* muG
E2G <- (0.15^2*(2^2) + 2*(0.15)*(0.85)*(1^2) + 0.85^2*(0^2) )
VarG <- (E2G-muG^2)
VarY= 0.4^2 *(E2G-muG^2)+ 0.02 + 0.09
#VarY <- siga/(0.8)
H2 <- 0.16*VarG/VarY

sigd <- 0
#sigd <- VarG - siga
h2 <-siga/VarY
```