

LAB 1

1. Mean and variance, discrete distribution.

Assume the following discrete distribution function.

Genotype	y	$p(y)$
A_1A_1	5	$\frac{1}{4}$
A_1A_2	1	$\frac{1}{16}$
A_1A_3	2	$\frac{3}{16}$
A_2A_2	4	$\frac{1}{8}$
A_2A_3	3	$\frac{1}{8}$
A_3A_3	0	$\frac{1}{4}$

Calculate the following quantities.

- (a) $Pr(Y \leq 2)$
- (b) $Pr(0 < Y \leq 4)$
- (c) $E(y)$
- (d) $E(y^2)$
- (e) $Var(y)$

2. Unbiased Estimation of Variances

This is an example of Henderson's Method 1 of unbiased estimation of variance components. Let

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e},$$

with data as follows:

$$\begin{pmatrix} 15 \\ 42 \\ 20 \\ 36 \\ 50 \\ 17 \\ 34 \\ 23 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{12} \\ u_{13} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{21} \\ u_{22} \\ u_{23} \\ u_{24} \end{pmatrix} + \mathbf{e}.$$

Also,

$$\mathbf{V} = \mathbf{Z}_1\mathbf{Z}'_1\sigma_1^2 + \mathbf{Z}_2\mathbf{Z}'_2\sigma_2^2 + \mathbf{I}\sigma_0^2.$$

Calculate the following matrices:

- $\mathbf{M} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$
- $\mathbf{A} = \mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1$
- $\mathbf{B} = \mathbf{Z}_2(\mathbf{Z}'_2\mathbf{Z}_2)^{-1}\mathbf{Z}'_2$
- $\mathbf{Q}_0 = \mathbf{I} - \mathbf{M}$
- $\mathbf{Q}_1 = \mathbf{A} - \mathbf{M}$
- $\mathbf{Q}_2 = \mathbf{B} - \mathbf{M}$
- $\mathbf{y}'\mathbf{Q}_0\mathbf{y}$
- $\mathbf{y}'\mathbf{Q}_1\mathbf{y}$
- $\mathbf{y}'\mathbf{Q}_2\mathbf{y}$
- $E(\mathbf{y}'\mathbf{Q}_0\mathbf{y}) = \text{tr}(\mathbf{Q}_0\mathbf{V}_0)\sigma_0^2 + \text{tr}(\mathbf{Q}_0\mathbf{V}_1)\sigma_1^2 + \text{tr}(\mathbf{Q}_0\mathbf{V}_2)\sigma_2^2$
- $E(\mathbf{y}'\mathbf{Q}_1\mathbf{y})$
- $E(\mathbf{y}'\mathbf{Q}_2\mathbf{y})$
- Estimate the variances.
- Calculate the sampling variances.
- Calculate the sampling variance of the ratio of σ_A^2 to the sum of the three variances.
- What is the expectation of $\mathbf{y}'\mathbf{y}$ in general?

LAB 2

Likelihood Methods

Below are pedigrees and data on 14 animals, following a simple animal model.

Animal	Sire	Dam	Record
1	-	-	
2	-	-	
3	-	-	
4	-	-	
5	1	2	40
6	1	2	38
7	3	4	43
8	1	2	29
9	1	4	37
10	3	2	41
11	1	4	52
12	3	2	25
13	1	2	36
14	3	4	30

1. Construct \mathbf{A}^{-1} .
2. Set up the MME.
3. Apply EM-REML to the model,

$$y_{ij} = \mu + a_i + e_{ij}.$$

Let $\sigma_e^2/\sigma_a^2 = 1.5$ to start. Just do one or two iterations of EM-REML.

4. Apply DF-REML to the model in question 3. Try values of σ_e^2/σ_a^2 of 1.0, 1.5, 2.0, and 2.5, and determine the value that maximizes the log likelihood through quadratic regression.

LAB 3

Pedigrees and Single Trait Analysis

1. Prepare a pedigree file for analyses.

pedf03.f This program expects the pedigree file to contain the following information: animal ID, sire ID, dam ID, and birthdate(year). There should be one blank space between each variable. All data are expected to be numeric (i.e. no characters). Data are provided.

Sort the output file from pedf03.f according to the first 11 digits.

pedf04.f This program re-numbers animals consecutively, such that the new ID of a parent is less than any ID of its progeny. This structure is required for VCE and DMU too.

Sort the output file from pedf04.f again.

refrm.f This program uses the sorted output from pedf03.f in order to re-number the animal IDs in the data file. Thus, every animal ID in the data had to be in the pedigree file.

Sort the new data file by animal ID.

inbrd.f This program computes inbreeding coefficients for all animals in the pedigree.

2. Analyze the data using **pamst1.f** and **pamst2.f**. Use **plotg.sas** to determine the length of burn-in period, and **autoc.f** to determine the number of independent samples.

LAB 4

Software Tryouts

A fairly large dataset (around 14,000 records) and pedigree are available. There are two traits on each animal. The variables in the data file are

ID Herd ContG Sex Age Dam T1 T2

where **Herd** is the herd of birth, but **ContG** is the contemporary group effect, **Sex** is the sex of the animal, **Age** is the age of the animal when measured, **Dam** is for a maternal effect associated with trait 1, but not with trait 2, and **T1** and **T2** are the two traits.

The pedigree file has

Animal Sire Dam Birthyear

All animals are already numbered consecutively and parents have lower IDs than their progeny. There could be some inbreeding. Pick an appropriate model for each trait.

1. Prepare and run DMU1 and DMUAI for the two traits.
2. Run the data through VCE.
3. Run the data through MTDFREML.
4. Compare results from each software package in a table.

