**Lab 1 – BLUPF90 Family**

*Variance components, breeding values, and reliabilities*

Prepared by I. Aguilar, D. Lourenco, and S. Tsuruta

1. Documentation for BLUPF90 program in the wiki: <http://nce.ads.uga.edu/wiki/doku.php> and also in blupf90.pdf file.
2. Using the following example (from Mrode and Thompson, 2005 -*Linear Models for Predicting Animal Breeding Value*)

Create data, pedigree and parameter to run **renumf90** and then run **blupf90** to obtain solutions and reliabilities.

Reliability for animal i can be calculated as:

Reli = 1 – PEVi/VarA

Where:

PEV is the prediction error variance (S.E. = sqrt(PEV))

VarA is the additive genetic variance

**

The solutions from the example are:



Variance components estimation

1. Read documentation for REMLF90 program in the wiki: <http://nce.ads.uga.edu/wiki/doku.php> and also in remlf90.pdf file.
2. Files for this exercise are in directory day1/lab1

README file contains information about the data content. This is a real dataset used for evaluation of growth traits in beef cattle.

1. Create a renumf90 parameter file for YW (yearling weight) that fits the following model:

Y = CG + β AOD + animal + maternal + sire-herd + e

Consider CG as fixed; AOD as covariable; animal, maternal, sire-herd as random. Starting values for variance components are: $σ\_{a}^{2}=438.90$; $σ\_{m}^{2}=73.24$; $σ\_{am}=-35.80$; $σ\_{sh}^{2}=242.12$; $σ\_{e}^{2}=751.13$

Use depth of pedigree equal 3

1. Run renumf90 and check the output files.
2. Run airemlf90 with an option to get SE for heritability and for genetic correlation between direct and maternal effects.
3. Change the parameter file for renumf90 for a 3-trait model (BW, WW, YW) and run blupf90. Variance components are in the README file.
4. Run gibbs2f90 for the single trait example; use the number of samples 1000 and burn-in 0. Run postgibbsf90 with burn-in of 0; try burn-in of 200. For graphical output, postgibbsf90 requires a plotting package GNUPLOT and X Windows (e.g., as provided by X emulation packages: Xming).
5. Estimate breeding values using gibbs2f90. Initial values should be replaced by posterior means. Check the manual to find the correct option to get EBV and SE. Use 1000 samples and burn-in of 200.

**Lab2 – Single-Step GBLUP**

*Animal model for single-step GBLUP*

Prepared by D. Lourenco and I. Aguilar

The data for this lab was created using simulation for a single-trait animal model. Simulation was done using QMSim (Sargolzaei, M. and F. S. Schenkel. 2009)

The simulated phenotypes were generated using the following model:

*Phenotype = mean + true\_ebv + residual*

**Parameters**:

Mean = 1.0

True variances:

 direct genetic = 0.25

 residual =0.72

**Description of files**

**data.txt :**

1: Animal ID

2: Sire ID

3: Dam ID

4: Generation number

5: Sex code

6: Number of male progeny

7: Number of female progeny

8: Inbreeding coefficient

9: Homozogosity

10: Phenotype

11: Simulated residual

12: Polygenic effect

13: True EBV

14: Internal EBV from QMSim

15: Mean (column of ones to fit mean effect in BLUPF90)

**ped.txt :**

1: animal ID

2: sire ID

3: dam ID

**snps.txt :**

1: animal ID

2: marker information

1. Copy the full folder into your directory

cp –r day2/lab2 .

1. From raw data, modify renumf90 parameter file (renlab.par) according to the data file and to fit the following model for genomic evaluation:

*y = mean + animal + e*

1. Run renumf90 program to renumber data, pedigree file, and marker data.
2. Check the renf90.par, renf90.dat, and renaddxx.ped. From the renaddxx.ped file, identify genotyped animals, and check with wiki (<http://nce.ads.uga.edu/wiki/doku.php?id=readme.renumf90>) the content of each column. Print only information from genotyped animals into another file.
3. Estimate variance components considering and ignoring marker information.

From the airemlf90 output find the following statistics: number of genotyped animals, number of SNP markers

1. Run blupf90 without marker information using estimated variance components. Now run blupf90 using genomic information and compare cpu time and solutions. Obs: Check wiki to see how to read an external or pre-computed genomic matrix.
2. Validation on young candidates (individuals from 10th generation with no phenotypes).
3. Remove the phenotypic information from the 10th generation and obtain solutions from a model with marker information and with no marker information.

Hint: if generation information is in column 4, the new data can be created using the AWK Linux tool

awk ‘$4!=10’ renf90.dat > renf90.dat.pred

1. Compare correlations with true breeding values for genetic additive direct effect. Hint: have renumf90 passing to the data a column containing the generation information.
2. Matching G and A22.

Run blupf90 with the truncated data set (renf90.dat.pred) with the following option: OPTION tunedG 0

Compare statistics from G and correlations between G and A22 using OPTION tunedG 0 and the default (OPTION tunedG 2).

Check the average of EBV from both runs and compare them with the average of TEBV.

**OPTIONAL**

Exclude chromosomes with even numbers and repeat validation using truncated data and compare correlation with TEBV for reduced and full genome.

**Lab3 – Single-Step GWAS**

*Using different weights for SNPs*

Prepared by I. Aguilar, D. Lourenco, and H. Wang

The data for this lab was created using simulation for a single trait animal model. Simulation was done using QMSim (Sargolzaei, M. and F. S. Schenkel. 2009)

The simulated phenotypes were generated using the following model:

*Phenotype = mean + true-ebv + residual*

Files are available in the folder:

day2/lab3

**Description of files**

**pheno.txt:**

 1: mean

 2: animal id

 3: sire id

 4: dam id

 5: sex

 6: generation

 7: number of male progenies

 8: number of female progenies

 9: inbreeding

10: homozygosity

11: phenotype

12: simulated residual (e)

13: individual true breeding value for polygene

14: individual true breeding value for direct effect (qtl)

15: EBV from QMSim internal BLUP

**pedigree.txt :**

1: animal ID

2: sire ID

3: dam ID

**mkr.txt :**

1: animal ID

2: marker information

**chrmap**:

1: SNP ID

2: Chromosome

3: position

1. Copy the full folder into your directory

**cp –r day3/lab3 .**

Go to <http://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90>

 and check all options available for postGSf90

1. Run renumf90 program using ‘renum.par’ parameter file to renumber data, pedigree, and marker files
2. Run blupf90 and get solutions

1. Add an option to read a map file (*chrmap*) and run postGSf90. Check the output files.

1. Try analyses using option to get variance explained by windows of adjacent SNPs and add the option to generate Manhattan plots. Check the output files.
2. Prediction of DGV for young individuals. The postGSf90 creates a file *snp\_pred* with information about the random effect (number of traits + correlated effects), the gene frequencies and the solutions of SNP effects.

Use the program predf90 to predict DGV using a marker file for young individuals (*young\_anim*)