File S11

Script S1: SAS PROC MIXED for Polygenic Variance Components Analysis

This is the program code for PROC MIXED in SAS. The code also includes PROC IMPORT used to read the input data. The outputs are directly printed out in the window. In addition, estimated parameters and predicted genomic values are also written in SAS datasets. These SAS datasets can be exported later into physical files using PROC EXPORT (not provided). To make sure that PROC MIXED produces legal estimates of variance components, a lowerb= option is given in the parms statement. The lower bound option of 1e-5 means that each variance component is bounded at 1e-5, i.e., $\sigma_X^2 \geq 1e-5$. There are seven estimated variance components (six polygenic variances and one residual variance). All initial values of the variances are set to 1.0. Users can choose different initial values, depending on the properties of the data. The initial value of one (1) is the default initial value for the variance parameters in PROC MIXED. The trait shown in the code is KGW.

```
/*begin code*/

%let dir=C:\Users\SHXU\Programs;
filename kk "&dir\Data S1.csv" lrecl=20000;
filename phe "&dir\Data S2.csv";

proc import datafile=kk out=kk dbms=csv replace;
proc import datafile=phe out=phe dbms=csv replace;
run;

proc mixed data=phe method=reml;
class line;
model kgw=/solution;
random line/type=lin(6) ldata=kk solution;
parms (1) (1) (1) (1) (1) (1) (1)/lowerb=1e-5 1e-5 1e-5 1e-5 1e-5 1e-5 1e-5;
ods output SolutionR=blup SolutionF=fixed CovParms=covar;
run;

data pred;
   merge phe blup;
run;

proc corr data=pred;
   var kgw estimate;
run;

/*end code*/
```

Comments: The program takes two input files stored in a user defined folder (c:\users\shxu\programs in this example), one file for the kinship matrices (named Data S1.csv in this example) and one for the phenotypic values (named Data S2.csv in this example). The Data S2.csv file must contain a variable for the id number of lines (named line in this example) and a variable for the phenotypic values of the trait in question (named kgw in this example). The program will generate three SAS datasets. One SAS dataset is called blup, which gives the predicted polygenic value for each line, the second SAS dataset is named fixed, which gives the estimated fixed effects and the third SAS dataset named covar gives all the seven estimated variance components, including six polygenic variances and one residual variance. The two input files are provided in Supplemental Data S1 for the kinship matrix and Data S2 for fixed-effect adjusted phenotypic values of 278 lines for four quantitative traits.