File S12

Script S2: SAS PROC HPMIXED for Association Studies

This is the program code for PROC HPMIXED in SAS. The code also includes PROC IML for eigenvalue/eigenvector calculation and data manipulation. This program only shows the mixed model association study for two bins, bin1 = 729 and bin2 = 1064. The trait analyzed is KGW because this trait has shown that the two bins have strong interactions in the whole genome analysis. The model contains eight genetic effects (a1, a2, d1, d2, aa, ad, da, and dd). The program requires polygenic variance ratios (lambda values) calculated from the PROC MIXED program. The SAS dataset named lambda contains pre-calculated lambda values for all the four traits and thus the data matrix dimension is 6×4 (six observations and four variables). The program will print all outputs on the screen and also write various output tables into SAS datasets. The most important output is the estimated genetic effects in an output dataset called blup1. In the script, the PROC HPMIXED program is called twice, one for the polygenic model (null model) without fitting the two bins. This call produces a likelihood value ($-2L_0$) under the null model. The second call of this procedure produces a likelihood value ($-2L_1$) under the full model (fitting the two bins). A dataset called lrt is generated by taking the difference between the two likelihood values, lrt = ($-2L_0$) – ($-2L_1$) = $-2(L_0-L_1)$. This likelihood ratio test statistic is testing the significance of the two bins jointly.

```
/*begin code*/

%let dir=C:\Users\SHXU\Programs;
filename kk "&dir\Data S1.csv" lrecl=20000;
filename phe "&dir\Data S2.csv";
filename gen "&dir\Data S3.csv" lrecl=20000;

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

data lambda;
   input yield tiller grain kgw;
cards;
5.09424E-06   1.190743713   7.00747898    20.04424779
5.09424E-06   2.67523E-05   4.64533E-07   8.84956E-05
3.570809985   1.563937935   3.10387885    2.761061947
2.685277636   2.67523E-05   0.305281739   2.010619469
9.310901681   0.322632424   5.118688159   8.84956E-05
2.826439124   3.024879615   4.64533E-07   1.666371681
;

proc iml;
use lambda;
   read all var{kgw} into lambda;
close;
use phe;
   read all var{kgw} into y;
close;
p=nrow(lambda);
n=nrow(y);
kk=j(n,n,0);
do k=1 to p;
   range=((k-1)*n+1):(k*n);
   use kk;
      read point range into kk0;
   close;
   kk0=kk0[,3:(n+2)];
   kk=kk+kk0*lambda[k];
end;
call eigen(delta,uu,kk);
create delta from delta;
   append from delta;
```

```
close;
create uu from uu;
   append from uu;
close;
x=j(n,1,1);
xu=uu`*x;
yu=uu`*y;
w=1/(delta+1);
yxw=yu||xu||w;
varname={"y" "x" "w"};
create yxw from yxw[colname=varname];
   append from yxw;
close;
quit;

proc hpmixed data=yxw;
   model y = x/noint solution;
   weight w;
   ods output CovParms=parm0 FitStatistics=fit0 ParameterEstimates=fixed0;
   nloptions  maxiter=10000 gconv=1e-8;
run;

proc iml;
use gen;
  read all var{bin729 bin1064} into zz;
close;
k=1;
l=2;
zk=(zz[,k]='A')-(zz[,k]='B');
wk=(zz[,k]='H');
zl=(zz[,l]='A')-(zz[,l]='B');
wl=(zz[,l]='H');
create zz from zz;
append from zz;
close;

use yxw;
  read all into yxw;
close;
use uu;
   read all into uu;
close;
z=zk||zl||wk||wl||(zk#zl)||(zk#wl)||(wk#zl)||(wk#wl);
zu=uu`*z;
yxwz=yxw||zu;
varname={"y" "x" "w" "a1" "a2" "d1" "d2" "aa" "ad" "da" "dd"};
create yxwz from yxwz[colname=varname];
    append from yxwz;
close;
quit;

proc hpmixed data=yxwz;
   effect z=collection(a1 a2 d1 d2 aa ad da dd);
   model y=x/noint solution;
   weight w;
   random z / solution;
   ods output CovParms=parm1 FitStatistics=fit1
              ParameterEstimates=fixed1
              SolutionR=blup1 ConvergenceStatus=conv1;
   nloptions  maxiter=10000 gconv=1e-8;
run;

data lrt;
```

```
   merge fit0(rename=(value=l0)) fit1(rename=(value=l1));
   lrt=l0-l1;
   if _n_=1;
run;

/*end code*/
```

One of the outputs generated from PROC HPMIXED is the estimated genetic effects for the two bins (bin1 = 729 and bin2 = 1064).

Solution for Random Effects

Effect	z	Estimate	Std Err Pred	DF	t Value	Pr > |t|
z	a1	0.5388	0.1688	277	3.19	0.0016
z	a2	0.09325	0.1619	277	0.58	0.5651
z	d1	-0.1718	0.1315	277	-1.31	0.1924
z	d2	-0.07827	0.1302	277	-0.60	0.5482
z	aa	0.2817	0.1164	277	2.42	0.0162
z	ad	0.3970	0.1296	277	3.06	0.0024
z	da	0.08289	0.1327	277	0.62	0.5329
z	dd	-0.1170	0.1595	277	-0.73	0.4636

Comments: The program takes three 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), one for the phenotypic values (named Data S2.csv in this example) and the third file for the bin genotypes (named Data S3 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 also requires a SAS dataset named lambda to store the six variance ratios obtained from the polygenic analysis via PROC MIXED described in Script S1. The program will generate several SAS datasets. One SAS dataset is called blup1, which gives the estimated genetic effects in the following order: a1, a2, d1, d2, aa, ad, da and dd. The three input files are provided in Supplemental Datas S1, S2 and S3, respectively.