Assignment 13

Ivan Pocrnic, 23.V.2018.

Please note that the code provided is partial

Assignment 13: 1, 2, 3, 4

Create: vi renum.par

```
#renumf90 parametar file
DATAFILE
data3.txt
TRAITS
4
#Phen
FIELDS_PASSED TO OUTPUT
1 5 2
#ID TBV GEN
WEIGHT(S)

RESIDUAL_VARIANCE
0.6
EFFECT
3 cross alpha
#sex
EFFECT
1 cross alpha
#animal
RANDOM
animal
OPTIONAL

FILE
ped3.txt
SNP_FILE
snp3.2k
(CO)VARIANCES
0.4
```

Run RENUM:

```
echo renum.par | renumf90 | tee renum.log
```

Identify genotyped animals:
Please note this can be done in several ways

```
awk '{if ($6>9) print $0}' renadd02.ped | sort +0 -1 > gen.ped
wc -l gen.ped
#2024
wc -l snp3.2k
#2024
```

Assignment 13: 5

Run PREGS 1st time:

```
cp renf90.par pregs1.temp
echo "OPTION saveCleanSNPs" >> pregs1.temp
echo "OPTION chrinfo mrkmap.txt" >> pregs1.temp
echo pregs1.temp | preGSf90 | tee pregs1.log
```

Assignment 13: 6

Run PREGS 2nd time, using clean file!

```
cp renf90.par pregs2.temp
sed -i 's:snp3.2k:snp3.2k_clean:g' pregs2.temp
echo "OPTION plotpca" >> pregs2.temp
echo "OPTION chrinfo mrkmap.txt_clean" >> pregs2.temp
echo pregs2.temp | preGSf90 | tee pregs2.log
```

Assignment 13: 7

Run BLUP with full data:

```
mkdir blup; cd blup
cp ../renf90.par par.temp
grep -v OPTION par.temp > blup.par
sed -i 's:renf90.dat:../renf90.dat:g' blup.par
sed -i 's:renadd02.ped:../renadd02.ped:g' blup.par
time (echo blup.par | blupf90 | tee blup.log)
cp solutions blup_solutions
rm *.temp
```

```
cd ..
```

Run ssGBLUP with full data:

```
mkdir gblup; cd gblup
cp ../renf90.par par.temp
grep -v OPTION par.temp > gblup.par
sed -i 's:renf90.dat:../renf90.dat:g' gblup.par
sed -i 's:renadd02.ped:../renadd02.ped:g' gblup.par
echo "OPTION SNP_file ../snp3.2k_clean" >> gblup.par
echo "OPTION chrinfo ../mrkmap.txt_clean" >> gblup.par
echo "OPTION no_quality_control" >> gblup.par
time (echo gblup.par | blupf90 | tee gblup.log)
cp solutions gblup_solutions
rm *.temp
cd ..
```

We are using clean file, but keep in mind same result if just use blupf90 directly

Assignment 13: 8

Run Validation:

Create reduced dataset:

```
awk '$6!=5' ../renf90.dat > renf90.dat.reduced
```

Now repeat as in 13.7 - run BLUP & ssGBLUP but this time with reduced data (renf90.dat.reduced). After running, prepare validation files.

For example, validation is on animals from generation 5 that had phenotypes removed in reduced, and had genotypes.

```
awk '$6==5' ../renf90.dat | sort +2 -3 > young.dat
awk '{print $1}' ../snp3.2k_clean_XrefID | sort +0 -1 > gen.ped
```

Merge created files togenther with your solution files from BLUP reduced and ssGBLUP reduced. After that use your favourite stat software for calculating results.

For example in R, if in the column 3 is TBV and column 4 EBV:

```
a=read.table(file="solucije")
cor(a$V3,a$V4)
reg=lm(a$V3~a$V4)
summary(reg)
```

Repeat the same with GEBV.

Assignment 13: 9

Predictivity / predictive ability,go to the existing BLUP folder where you have the results from BLUP with full/complete data.

```
echo 'OPTION include_effects 2' >> pred.par
echo pred.par | predictf90 | tee pred.log
```

Take the results and extract y-Xb

```
awk '{print $1,$2}' yhat_residual | sort +0 -1 > yxb.tmp
```

Merge this file with your solutions and do the validation using your favourite stat software. This time you can use the formula Acc=cor(y-Xb,GEBV)/sqrt(h2).

Assignment 13: 10

We already have (G)EBV from full and from reduced data - no aditional calculations needed! We will valiadate on the same number of young animals from gen 5 as in previous! So you only need to merge all the data together and use your favourite stat software to get the results (All the formulae are provided in the assignment).