Day 4 Genomic Prediction and Genomic Selection

Exercise 1 Accuracy of Genomic Predictions

Use the spreadsheet '*GSaccuracy 2015.xls'* to investigate and explain the impact of the following parameters on the effective number of chromosome segments (M_e), the proportion of variance explained by markers (q^2), the accuracy with which marker effects are estimated ($r_Q$hat), and the accuracy of the genomic prediction or molecular breeding value, MBV (r_{MBV}):

- number of markers (M)
- Effective population size (N_e)
- Heritability of phenotypes (h^2)
- Number of training individuals (N)

Set L=1 and k=30 for a genome of 30 chromosomes of 1 Morgan

a) What is the minimum number of markers that is needed to achieve near maximum genome coverage (q^2=0.99) when N_e=100 versus 340 versus 1000 versus 10,000? Enter this in the table below.

b) Set the number of markers M = 1,000,000 to get nearly complete coverage regardless of N_e. Set h^2=0.9.
Now evaluate the size of the training set (N) needed to reach an MBV accuracy of 0.8 for N_e=100, versus 340 versus 1,000 versus 10,000. Enter the results in the table below.

c) Repeat b) for heritabilities equal to 0.5 and 0.2

	N_e = 100	**N_e = 340**	**N_e = 1000**	**N_e = 10,000**
Min. # markers	**125k**	**350k**	**900k**	**6,5M**
h^2 = 0.9	**2,600**	**7,100**	**17,800**	**166,000**
h^2 = 0.5	**4,700**	**12,650**	**32,000**	**300,000**
h^2 = 0.2	11,700	31,600	80,000	750,000

d) Test the genome scaling argument that if the size of the simulated genome is reduced by a factor C, then the size of the training population also has to be reduced by the same factor C in order to maintain the same accuracy of MBV.

Exercise 3.2 EBVs combining phenotype and GBV

Consider selection for a trait with (total) heritability 0.2 and phenotypic standard deviation equal to 50. For this trait, you have been able to derive genomic breeding values with accuracy (r_{MBV}) equal to 0.5. The following sources of information are available for selection:

- own MBV
- own phenotype
- MBV of the individual's sire
- Phenotype of the individual's sire

Assume the top 20% individuals are selected. The base scenario is having own phenotype and phenotype of the sire only (no genotypes). Evaluate the impact of the following on the accuracy of selection and the genetic superiority of the selected individuals, using the spreadsheet 'STEBVaccuracy.xls':

a) Adding genotyping of the individual to the base information of own phenotype and phenotype of the sire.

b) Adding genotyping of the sire to the information from question a).

c) Derive the elements of the P matrix and the G vector for question b)

d) Explain the weight on the MBV of the sire in question b)

e) Convert this MBV blending problem into a two-trait problem consisting of the original trait and the MBV as a correlated trait with heritability = 1. I.e. derive the genetic and phenotypic correlations between the two traits. Enter the results into MTindex.xls and check that it gives the same answers for question b).

Base: no genotyping

Parameters	
Heritability	0.2
Repeatability of subsequent records	0.4
c-squared (among full sibs)	0.15
Reliability of genomic test (prop expl)	0.09

Run

Information used	Nr.Records	Genotyped	Index Wght		Value of Variate	
			Pheno	Geno	Pheno	Geno
nr of own records	1	0	0.192	-	53.6%	-
nr. of records on dam	0	0	-	-	-	-
nr of records on sire	1	0	0.081	-	7.2%	-
nr of fulls sib records	0	0	-	-	-	-
nr. of half sib records (excl. full sibs)	0	0	-	-	-	-
nr. of progeny	0	0	-	-	-	-

Accurcay of EBV	**0.482**
Accuracy of Pheno EBV	0.482
Accuracy of MBV	0.000
correlation EBV FS	0.405
correlation EBV HS	0.247

50 PhenSD
22.36068 GenSD
15.08896 superiority top 20%

a)

Parameters	
Heritability	0.2
Repeatability of subsequent records	0.4
c-squared (among full sibs)	0.15
Reliability of genomic test (prop expl)	0.09

Run

Information used	Nr.Records	Genotyped	Index Wght Pheno	Index Wght Geno	Value of Variate Pheno	Value of Variate Geno
nr of own records	1	1	0.178	0.784	32.2%	9.9%
nr. of records on dam	0	0	-	-	-	-
nr of records on sire	1	0	0.075	-	5.0%	-
nr of fulls sib records	0	0	-	-	-	-
nr. of half sib records (excl. full sibs)	0	0	-	-	-	-
nr. of progeny	0	0	-	-	-	-

Accurcay of EBV	**0.535**
Accuracy of Pheno EBV	0.482
Accuracy of MBV	0.300
correlation EBV FS	0.443
correlation EBV HS	0.262

50 PhenSD
22.36068 GenSD
16.75593 superiority top 20%

Increase in accuracy = 11%

b)

Parameters	
Heritability	0.2
Repeatability of subsequent records	0.4
c-squared (among full sibs)	0.15
Reliability of genomic test (prop expl)	0.09

Run

Information used	Nr.Records	Genotyped	Index Wght Pheno	Index Wght Geno	Value of Variate Pheno	Value of Variate Geno
nr of own records	1	1	0.178	0.822	32.1%	8.2%
nr. of records on dam	0	0	-	-	-	-
nr of records on sire	1	1	0.076	-0.076	5.0%	0.1%
nr of fulls sib records	0	0	-	-	-	-
nr. of half sib records (excl. full sibs)	0	0	-	-	-	-
nr. of progeny	0	0	-	-	-	-

Accurcay of EBV	**0.536**
Accuracy of Pheno EBV	0.482
Accuracy of MBV	0.300
correlation EBV FS	0.433
correlation EBV HS	0.246

50 PhenSD
22.36068 GenSD
16.7672 superiority top 20%
15.088
1.1113

Accuracy is about the same as previous, as we don't get much information out of the sire's genotype if the animal has been genotyped already. The accuracy is increased very slightly because we can weigh the polygenic information from the sire more appropriately, by adjusting the phenotype for the genotyped effect. There are different (in fact opposite) weights for the phenotype of the sire and the genotype of the sire. So the index is constructed as EBV = w1 geno + w2 pheno + w3 (pheno sire – geno sire)

Exercise 3.3 Genomic Selection and crossbreeding

Consider a purebred line producing pigs for crossbred performance.
Use CROSSBRGS.XLS to work out selection accurcay of the Index.
The breeding goal is 100% CB information.
Selection in the purelines is based on own performance, sire, dam, 3FS and 40HS.
In addition here are 40 Xbred HS performances recorded.
Trait heritability 0.25; c-squared 0.15.

Work out the accuracy of the index for 2 values of r_{PC} (correlation between purebred and crossbred performance):

Cases	r_{PC} = 0.7	r_{PC} = 0.9

No Genomic Selection
No information from Xbreds
With HS Information from Xbreds

Genomic Selection (Ne=100)
With HS Information from Xbreds
RefPopSize = 3000 PB

Genomic Selection
With HS Information from Xbreds
RefPopSize = 1500PB and 1500XB

Genomic Selection
With HS Information from Xbreds
RefPopSize = 3000 XB

This exercise was not discussed in detail