

Practical Tuesday 20 October morning

Part 1. Standard error of the genetic correlation

Use for this exercise the R-code provided in the paper by Bijma and Bastiaansen. It is made for looping across different parameters. The program was made for the standard error of the genetic correlation between purebred and crossbred performance. For this practical, you can read purebred and crossbred as environment 1 and environment 2.

1. Make a plot of the standard error of the genetic correlation as a function of the heritability when the genetic correlation is 0.8, the number of half-sib families is 100 (one offspring per dam; 100 sires) and number of offspring per environment is 50. Ignore the common environmental effect ($c^2=0$).

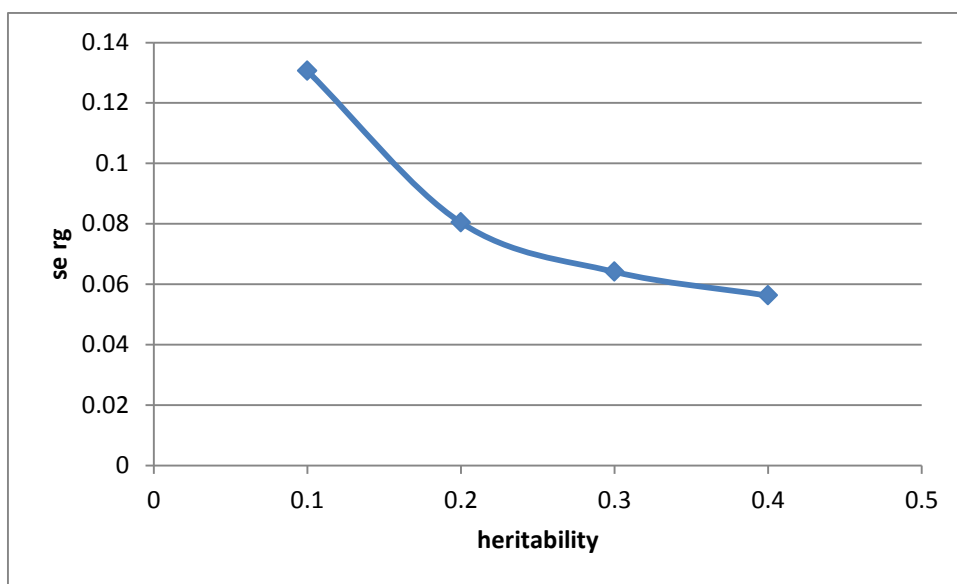


Figure 1. The effect of heritability on the standard error of the genetic correlation between two environments when having 50 half-sibs per sire per environment and having 100 sire families.

If the heritability increases, the standard error of the genetic correlation goes down (Figure 1). This means that the precision of estimating the genetic correlation is higher for traits with a high heritability.

2. Make a plot of the standard error of the genetic correlation as a function of the number of offspring per environment when the genetic correlation is 0.8, the number of half-sib families is 100 (one offspring per dam; 100 sires) and the heritability is 0.1, 0.3 or 0.5. Ignore the common environmental effect ($c^2=0$).

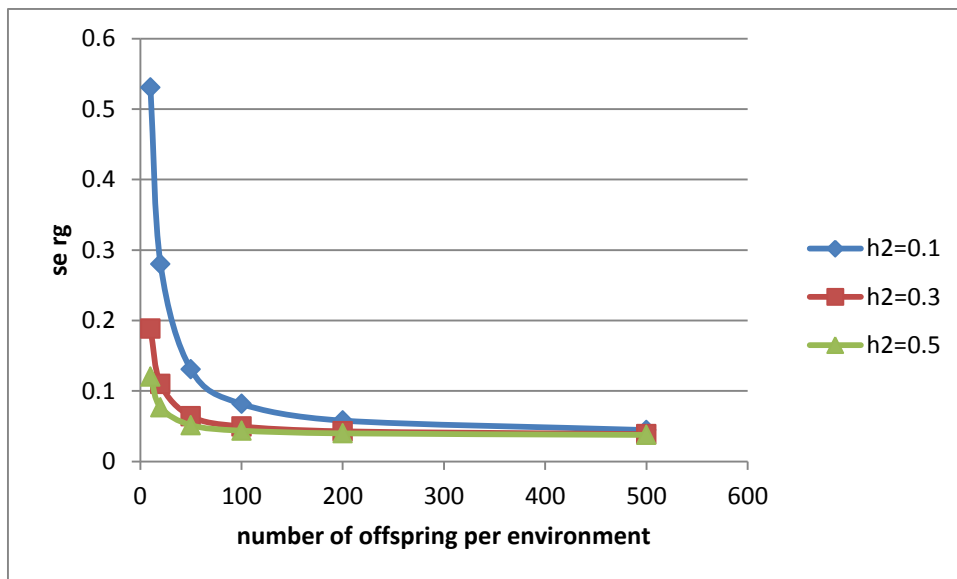


Figure 2. The effect of number of offspring per environment on the standard error of the genetic correlation between two environments when having 100 sire families for different heritabilities.

The standard error of the genetic correlation decreases with an increasing number of offspring per environment (Figure 2). If the heritability low, a larger number of offspring per environment is needed to get an estimate of a genetic correlation with a reasonable standard. For instance, to get a standard error below 0.1, one needs ~80 offspring per environment.

3. Make a plot of the standard error of the genetic correlation as a function of the number of families when the genetic correlation is 0.8, the number of offspring per environment is 50 and the heritability is 0.1, 0.3 and 0.5. Ignore the common environmental effect ($c^2=0$).

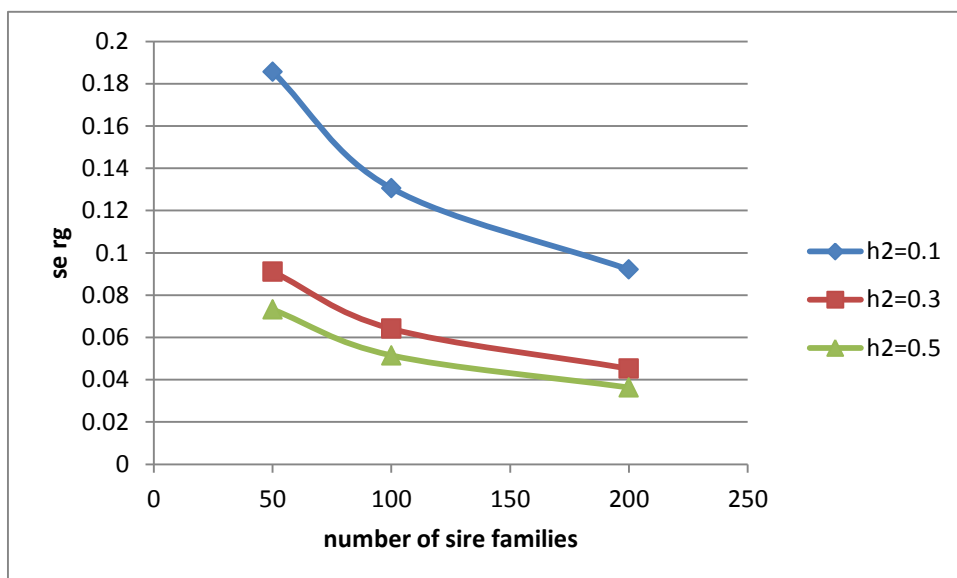


Figure 3. The effect of number of sire families on the standard error of the genetic correlation between two environments when having 50 offspring per family per environment for different heritabilities.

Figure 3 shows that more families are needed when the heritability is 0.1 than when it is 0.3 or 0.5 to get the same standard error on the genetic correlation.

4. Let's consider a trait with quite large common environmental effects such as harvest weight in fish. Typically, because of their small size as fingerlings, families are kept together for quite some time in family tanks. Make a plot of the standard error of the genetic correlation as a function of c^2 , the ratio of common environmental variance to the phenotypic variance, when the genetic correlation is 0.8, the heritability is 0.3, the number of sires is 50 and the number of dams mated to each sire is 2. Each dam has 50 offspring in one environment.

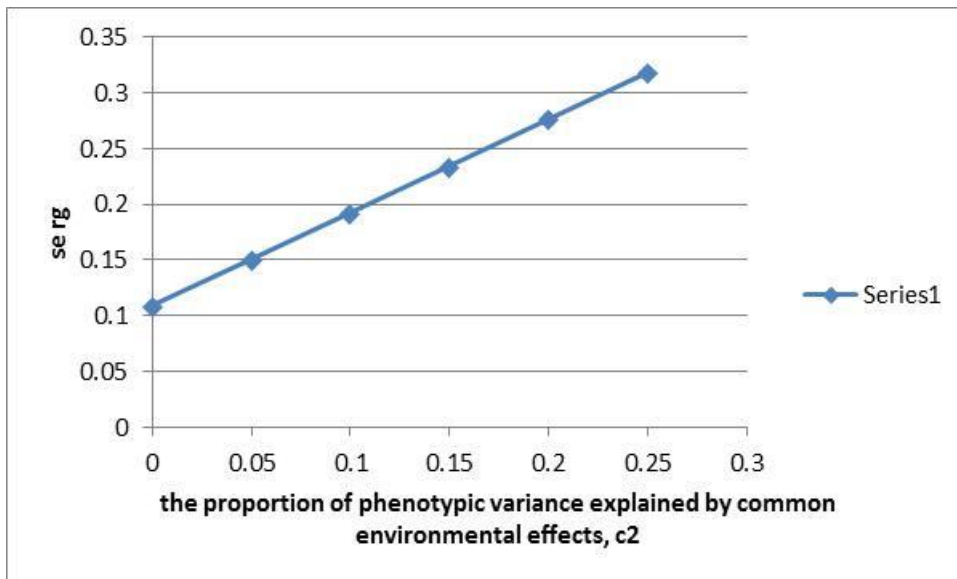


Figure 4. The effect of the common environmental variance (c^2) on the standard error of the genetic correlation between two environments when having 50 offspring per dam, 2 dams per sire, 100 sires and heritability is 0.3.

Figure 4 shows that the common environmental effect has a detrimental effect on the standard error of the genetic correlation. Therefore, it is desirable to keep families as short possible together to minimize c^2 .

5. (challenging) Adapt the R-code in which the total number of animals per environment is constant and vary the number of families and family size and determine whether there is an optimum. Vary the heritability. Ignore the common environmental effect ($c^2=0$). Set the number of animals per environment equal to 2000. Assume as in question 4 that each sire is mated to 2 dams for each environment.

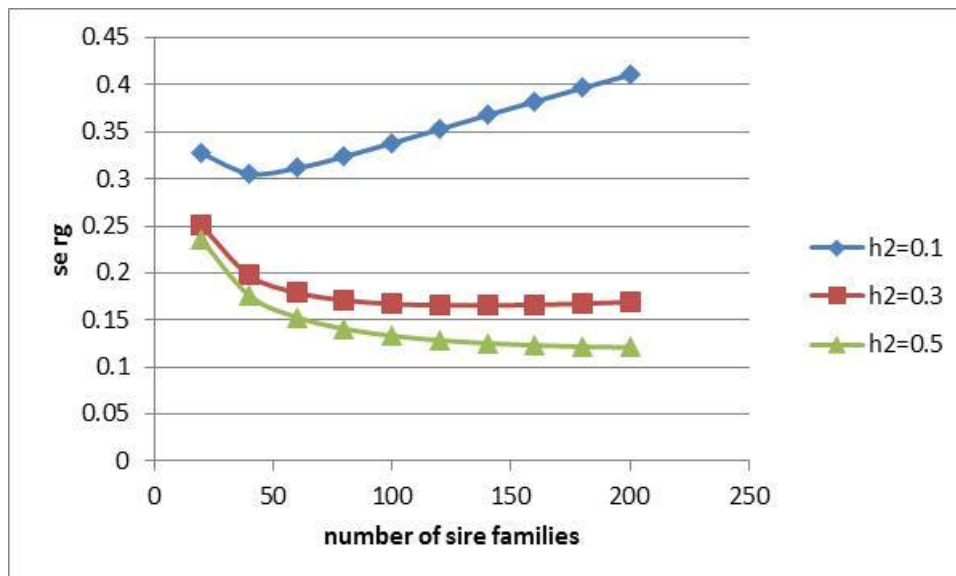


Figure 5. The effect of number of sire families on the standard error of the genetic correlation between two environments when the total number of animals per environment is 2 and when each sire is mated to 2 dams per environment for different heritabilities.

Figure 5 shows that there is an optimum number of sire families when the heritability is 0.1 leading to lowest standard error (optimum is about 40 sire families). When the heritability is 0.3, the optimum is 140 sire families and when the heritability is 0.5, the optimum is about 250 sire families. So clearly the optimum number of sire families increases when the heritability increases. In other words, with a low heritability it is important to have fewer families with more offspring per family, while with heritability it is better to have more families that are smaller in size.

Part 2. Bivariate and reaction norm models in ASREML

The aim of this exercise is to work with ASReml to estimate a genetic correlation between two environments in a bivariate analysis and to analyse a data set using a reaction norm model. In both cases, you will use simulated data. The focus is on interpreting the output from ASReml. For the bivariate model (bivariate.zip), the used simulated dataset cows.dat contains 10,000 cows, from which 4900 are in environment 1 and 5100 in environment 2. The 10,000 cows are daughters from 100 sires, each with 100 daughters. Each sire has approximately half of the daughters in each environment. For the reaction norm model (RN.zip), the used dataset is cows_asreml.dat. The dataset is slightly different than the previous one, but has the same structure with 100 sires each with 100 daughters.

1. Look at the biv.as file and the cows.dat file to understand the analysis.

The data is sorted on the environment and the first 4900 records belong to environment 1; the second 5100 belong to environment 2.

2. Run the biv.as and the pin-file in ASREMLW. Study the output in biv.asr and in biv.pvc. How do you interpret the genetic correlation and its standard error?

The variance-covariance matrix is:

Covariance/Variance/Correlation Matrix US x.animal
 0.2340 0.9243 # second value is correlation
 0.2763 0.3818 # first value is covariance

A genetic correlation of 0.9243 with se=0.0486 shows that there some reranking, but not a lot. In other words, the GxE causing reranking is small. When one looks at the genetic variance in the two environments 0.23 and 0.38, there is heterogeneity of variance.

3. Is the standard error according to your expectation, e.g. when comparing to the program se_rg.R?

The program se_rg gives an se of 0.0549, slightly higher than estimated by ASREML.

4. Open now the RN.zip and study the RN.as and the datafile cows_asreml.dat.

5. Run ASReml and run the pin-file. Study the output in RN.asr and RN.pvc. what is you conclusion with respect to GxE?

The variance-covariance matrix estimated by ASREML is:

0.2855 -0.1218 # variance in intercept and correlation intercept and slope
 -0.1942E-01 0.8907E-01 # covariance intercept slope, variance in slope

The pin-file instructs ASReml to calculate genetic variances per environment i using the formula:

$$var(A)_i = \sigma_{Aint}^2 + 2x_i\sigma_{Aint,Asl} + x_i^2\sigma_{asl}^2$$

The genetic covariance between environments is:

$$Cov(A_i, A_j) = \sigma_{Aint}^2 + (x_i + x_j)\sigma_{Aint,Asl} + x_i x_j \sigma_{asl}^2$$

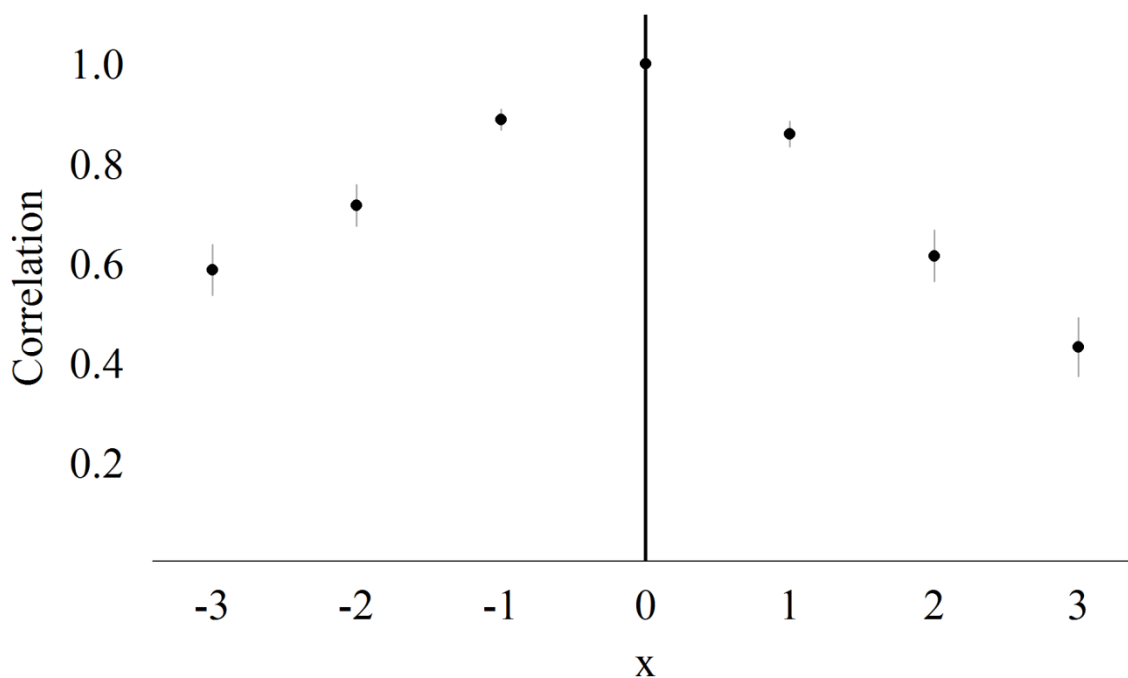
And then finally correlations are calculated using the covariances and the variances.

```
cor      = covar 2/SQR[var1 2 *var2 2 ]= 0.5864 0.0509
cor      = covar 2/SQR[var1 2 *var2 2 ]= 0.7156 0.0416
cor      = covar 2/SQR[var1 2 *var2 2 ]= 0.8875 0.0206
cor      = covar 2/SQR[var1 2 *var2 2 ]= 1.0000 -0.0000 Not estimable
cor      = covar 2/SQR[var1 2 *var2 2 ]= 0.8594 0.0253
cor      = covar 2/SQR[var1 2 *var2 2 ]= 0.6146 0.0510
cor      = covar 2/SQR[var1 2 *var2 2 ]= 0.4317 0.0589
```

Based on these correlations, we can conclude that there is substantial reranking due to GxE between extreme environments.

6. Plot the genetic correlation and its standard error between different pairs of environments using the provided r-code standard error rg.R.

The plot looks like:



7. Change the intercept of the reaction norm by subtracting/adding a certain value to x , e.g. (-1, -2, +1, +2). Write down the REML likelihood and the estimated variance components. What is your conclusion?

x+delta	logL	vara int	vara slope	cor int sl
0	-5301.61	0.285503	8.91E-02	-0.1218
1	-5301.61	0.413458	8.91E-02	-0.5654
2	-5301.61	0.719483	8.91E-02	-0.7804
3	-5301.61	1.2037	8.91E-02	-0.8754

If we add 1, 2 or 3 to the x -variable, we shift the intercept to the right. We observe no change in LogL and in the genetic variance in slope, but we do see changes in genetic variance in intercept and in the correlation between intercept and slope.

8. If time permits you may want to try Legendre polynomials by using `leg(x,1).animal`. You can see in the `.res` file the polynomial coefficients.

The output in `asr.file` is:

```
8 LogL=-5300.96   S2= 0.73592   9998 df
```

```
- - - Results from analysis of Pheno - - -
```

```
Akaike Information Criterion  10609.91 (assuming 4 parameters).
```

```
Bayesian Information Criterion  10638.75
```

Model_Term	Gamma	Sigma	Sigma/SE	% C
Residual	SCA_V 10000 1.000000	0.735921	19.08	0 P

```

leg(x,1).animal    US_V 1 1 0.889457    0.654570    6.87 0 U
leg(x,1).animal    US_C 2 1 -0.237492   -0.174776   -6.45 0 U
leg(x,1).animal    US_V 2 2 0.451003    0.331903    7.68 0 U
Covariance/Variance/Correlation Matrix US leg(x,1).animal
0.6546   -0.3750
-0.1748   0.3319

```

So the likelihood is almost the same as before. If we want to calculate the genetic variance for instance in environment $x=0$ than we get the polynomial coefficients 0.70711 0.25995 from the Res-file. This would mean that the genetic variance in environment $x=0$ is:

$$[0.70711 \quad 0.25995] \begin{bmatrix} 0.654570 & -0.174776 \\ -0.174776 & 0.331903 \end{bmatrix} \begin{bmatrix} 0.70711 \\ 0.25995 \end{bmatrix} = 0.285461$$

So this answer is basically the same as without legendre polynomials.