Cheat sheet on how to calculate *predicted* genetic gains

Excellence in Breeding Platform

Donors and stakeholders of breeding organizations expect to receive values for different terms of the breeders' equation every year in order to assess the progress of pipelines towards the goals of greater genetic gains and variety turnover in the target population of environments

This cheat sheet provides a suggested template and guidelines to calculate and report these metrics.

Optimizing breeding schemes

Giovanny Covarrubias-Pazaran
g.covarrubias@cgiar.org

08/11/2021

It is suggested to use a basic spreadsheet that records the following data for each trait and reporting date.

Glossary

STG 1 Stage 1: The first year a cohort of germplasm is tested for yield (in addition to other traits) in a multi-location field trial.

KPI Key performance indicator.

Basic details

Reporting Date	Center	Crop	Region	Continent	Pipeline name	Trait name	Trait unit	Year
dd/mm/yyyy	ACRONYM					e.g. Drought tolerance	e.g. score 1-10	2020

> Average breeding cycle time

Cycle time	Idealized cycle time
Age of parents	Time from crossing to recycling

> Selection heritability and accuracy

Heritability (H^2)			SE-TPE correlation		
STG 1	STG 2	STG 3	STG 1	STG 2	STG 3

> Selection intensity

Proportion advanced		
STG 1	STG 2	STG 3

> Genetic variance

Genetic variance (σ^2_g)		
STG 1	STG 2	STG 3

Average breeding cycle time

Genetic cycle time	Idealized cycle time
Mean age of parents at time of crossing	Time from crossing block to recycling

The 1st KPI refers to the average age of individuals entering STG1.

Example

Consider a dataset with stage 1 materials for years 2019 and 2020. Materials tested contain in their GID the year when they were produced in the crossing block, or maybe such information is available from other sources (e.g., selection history in the database):

Table 1. 2019 dataset with materials tested in Stage 1 of yield testing.

GID	**Year of the cross**
G_001_2016	2016
…	…
G_026_2015	2015

Table 2. 2020 dataset with materials tested in Stage 1 of yield testing.

GID	**Year of the cross**
G_104_2017	2017
…	…
G_026_2015	2015

The genetic cycle time will be calculated as the year of recycling (current year) minus the average year of origin of the materials tested in that year.

$$GenCycleTime_{2019} = 2019 - \frac{(2016 + \cdots + 2015)}{n_{individuals}}$$

Something similar is done for every year to be reported in a summary table.

For the idealized cycle time you only have to report the intended time from cross to cross if all parents in the crossing block where coming from the recycling strategy.

Troubleshooting

Is it necessary to obtain the age of each individual in the stage selected?

No, only for the materials we have information when they were generated or introduced to the program.

What is the age of a parent that is introduced (i.e., not a recycled parent) and does not have any detailed performance data?

The year of origin for this material will be regarded as the year when the material was introduced if not detailed information is known.

Broad sense heritability and correlation

Heritability (H^2)		
STG 1	STG 2	STG 3

Here, we calculate the heritability of trials. Report the average of entry-mean heritabilities (or reliability) of the different environments explored at each stage, and record in parentheses the number of environments used in the calculation.

Example

Assume that the STG1 trials in the year 2019 were evaluated in two environments with an augmented design and 10% checks. A diagonal mixed model is fitted for the trait of interest.

In ASReml-R v4 language this is written as:

```
Fixed = Trait ~ Env

Random = ~diag(Env):Geno   OR
diag(Env):vm(Geno, A) # A is the
relationship matrix

Residual = ~dsum(~units | Env)
```

We then use the Cullis *et al*. (2006) formula:

$$H^2 = 1 - \frac{PEV_{\mu}}{2\sigma_g^2}$$

Where PEV is the predicted error variance and σ^2_g is the genetic variance in a pedigree-based, marker-based or simple genetic model. An R script for calculating H^2 is available at: gitlab.com/excellenceinbreeding/module2.

The genetic correlation of SE-TPE can be extracted from factor analytic models.

Troubleshooting

How is an average H^2 obtained?

Request the support from your Biometrician to calculate H^2, using a diagonal model to obtain a H^2 value for each environment where the STGn material was evaluated. The weighted average of these values is reported. It is also possible to use across-environment predictions to obtain an across-environment H^2.

Which H^2 formula should I use?

The Cullis *et al*. (2006) formula is used to calculate broad or narrow sense heritability with variance components coming from a diagonal model and then average. The variance component changes due to the availability of pedigree, marker or relationship data, and the value will be interpreted as narrow or broad sense heritability. Otherwise, use the regular entry-mean broad-sense heritability based on ratios of variance components. Note which method was used and whether kinship information was used.

Selection intensity

Proportion recycled/tested		
STG 1	STG 2	STG 3
20/1000	15/200	5/20

Provide, as a ratio, the number of parents recycled relative to the number of individuals tested in each stage (proportion recycled/tested column).

Example

Assume that STG1 contains 1000 individuals. First, we advance 20% of the best individuals from STG1 to STG2 (best 200), the best 10% from STG2 to STG3 (best 20) and so on: this can be seen as the "proportion advanced" (200/1000 and 20/200). Although the best 20% individuals from STG1 are moved to STG2, it is not certain that all will become parents. E.g., from the best 200 moved from STG1 to STG2, maybe 20 are used as parents in the crossing block, then the reported number should be 20/1000 (p=0.02).

The selection intensity can be easily calculated in R as `dnorm(qnorm(1 - p))/p` where p is the proportion of materials recycled.

Genetic variance

Genetic variance (σ^2_g)		
STG 1	STG 2	STG 3
20	15	10

This refers to the average genetic variance across the n environments where the material from STGx was evaluated (where x is a value equal or greater than one).

Example

Assume we evaluated our STG1 trials in two environments with an augmented design and 10% checks. We will fit a diagonal model for the trait of interest.

In ASReml-R v4 language that is:

```
Fixed = Trait ~ Env

Random = ~diag(Env):Geno   OR
diag(Env):vm(Geno, A) # A is the
relationship matrix

Residual = ~dsum(~units, Env)
```

We will then take the variance components for genotypes for each environment and average them to put it in the table.

References

Gustavo de Los Campos, Daniel Sorensen, and Daniel Gianola. "Genomic heritability: what is it?" PLoS Genet 11.5 (2015): e1005048.

Cullis, Brian R., Alison B. Smith, and Neil E. Coombes. "On the design of early generation variety trials with correlated data." Journal of Agricultural, Biological, and Environmental Statistics 11.4 (2006): 381-393.

excellenceinbreeding.org/toolbox/tools/eib-breeding-scheme-optimization-manuals

CGIAR | Excellence in Breeding Platform