Module 3: Heritability and Reliability Estimations in R

Fundamentals of Genomic Prediction and Data-Drive Crop Breeding

(November 24-28, 2025)

Waseem Hussain
Senior Scientist-I
International Rice Research Institute
Rice Breeding Innovations Platfrom
waseem.hussain@cgiar.org
whussain2.github.io

Mahender Anumalla
Scientist-I
International Rice Research Institute
South-Asian Hub, Hyderabad
m.anumalla@cgiar.org

Margaret Catolos
Associate Scientist
International Rice Research Institute
South-Asian Hub, Hyderabad
m.catolos@cgiar.org

November 20, 2025

Contents

Load the Libraries

```
> # Load the Required Libraries
>   rm(list=ls()) # Remove previous work
>   library(easypackages)
>   libraries("dplyr", "reshape2", "readxl", "lme4", "arm")
```

Upload the Phenotypic Data Sets

Here we will use the Rainfed Rice Breeding Trial data of IRRI which has 192 entries, 2 replications and 5 blocks and experimental design is Alpha lattice.

```
> # Read the saved csv file, if working directly
>   demo.data.filtered<-read.csv(file="demo.data.filtered2.csv",
+                                header = TRUE)
>   str(demo.data.filtered)
```

```
'data.frame':   1600 obs. of  11 variables:
 $ Environment: chr  "Env1" "Env1" "Env1" "Env1" ...
 $ Plot       : int  59 19 27 127 132 92 106 82 199 57 ...
 $ Genotype   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Rep        : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Block      : int  2 1 1 4 4 3 3 3 5 2 ...
 $ Row        : int  2 1 1 4 4 3 3 3 5 2 ...
 $ Column     : int  19 19 27 7 12 12 26 2 39 17 ...
 $ Line.type  : chr  "Entry" "Entry" "Entry" "Entry" ...
 $ DTF        : int  94 78 87 87 94 94 82 87 82 85 ...
```

```
 $ HT          : num  118 116 108 116 134 ...
 $ Yield       : num  5873 5844 5777 6284 5703 ...
```

```
> # factor conversion if below are not in factors
>   columns<-c("Environment", "Genotype", "Rep", "Block",
+              "Row", "Column", "Line.type")
>   demo.data.filtered[, columns]<-lapply(columns, function(x) as.factor(demo.data.filtered[[x]]))
>   demo.data.filtered$Yield<-as.numeric(demo.data.filtered$Yield)
>   demo.data.filtered$HT<-as.numeric(demo.data.filtered$HT)
>   demo.data.filtered$DTF<-as.numeric(demo.data.filtered$DTF)
> # Subset the required columns
>   demo.data.filtered<-demo.data.filtered[, c("Environment", "Genotype",
+                                "Rep", "Block", "Row","Column",
+                                "Line.type","Yield", "HT", "DTF")]
> # First we will arrange the rows and columns for spatial analysis.
> # Now we will subset the environments and Yields for analysis
>   demo.data.filtered<-data.frame(demo.data.filtered%>% group_by(Environment)%>%arrange(Row, Column))
>   demo.data.filtered<-data.frame(demo.data.filtered%>% arrange(Environment)) # Arrange by environment
```

Run the Model to Estimate Heritability

- Here in this section we will use open source R package called **lme4** to run the mixed-model and extract variance components
- More on this R package can be found here lme4 Tutorial 1, and lme4 Tutorial 2.

Heritability Based on Single Environment

- First let us subset the data for on environment to show how to perform the analysis for one trial or environment in lme4 R package
- We will run models which are feasible in lme4 R package.
- We will use basic models and show how to extract the results
- Subset the data for one environment first.

```
> # Subset the environment 1
>   sub.data<-subset(demo.data.filtered, Environment=="Env1")
>   sub.data<-droplevels.data.frame(sub.data)
```

Run the mixed model

Model 1.lme4

- The model described below is:

$y_{ijk} = \mu + g_i + r_j + b_{jk} + \epsilon_{ijk}$

$y_{ijk} =$ is the effect of ith genotype in jth replication and kth block within the jth replication

$\mu =$ overall mean

$g_i =$ random effect of the ith genotype

$r_j =$ fixed effect of the jth replication

b_{jk}= random effect of kth block nested within j replication

ε_{ijk} = residual error

here we assume errors are independent and identically distributed $\epsilon \sim iid\mathrm{N}(0, \sigma^2_\epsilon)$

```
> # Now apply model
>   model1<-lmer(Yield~Rep+(1|Genotype)+ (1|Rep:Block), data =sub.data)
```

Extract Results and Variance Components

- Here we will summarize the results using ***summary()*** function.
- The first few lines of output indicate that the model was fitted by REML as well as the value of the REML criterion.
- The second piece of the summary output provides information regarding the random-effects and residual variation.
- The third piece of the summary output provides information regarding the fixed-effects and the fourth piece of summary output provides information regarding the correlation of fixed effects.

```
> # Summarise the results
>   summary(model1)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: Yield ~ Rep + (1 | Genotype) + (1 | Rep:Block)
   Data: sub.data

REML criterion at convergence: 6239.3

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.90048 -0.59387  0.03899  0.60311  1.71001

Random effects:
 Groups    Name        Variance Std.Dev.
 Genotype  (Intercept) 431861   657.2
 Rep:Block (Intercept)  28499   168.8
 Residual              193255   439.6
Number of obs: 394, groups:  Genotype, 197; Rep:Block, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5233.19      94.20  55.552
Rep2           60.38     115.60   0.522

Correlation of Fixed Effects:
     (Intr)
Rep2 -0.614
```

Extract variance components

- Here we will extract variance components

```
>   Ve<- VarCorr(model1)
>   Ve
```

```
 Groups    Name        Std.Dev.
 Genotype  (Intercept) 657.16
```

```
Rep:Block (Intercept) 168.82
Residual              439.61
```

ANOVA for fixed effects

```
> # ANOVA
>   anova(model1)
```

	npar	Sum Sq	Mean Sq	F value
Rep	1	52724.26	52724.26	0.2728223

Extract the Fixed effects

- Here will show how to extract the BLUEs.

```
>   BLUEs<-fixef(model1)
>   BLUEs
```

```
(Intercept)        Rep2
  5233.1856     60.3794
```

Extract the Random effects

- Here will show how to extract the BLUPs.

```
> # Extract the Random effects
>   BLUPs<-data.frame(Blups.yield=ranef(model1)$Genotype)
>   GV<-data.frame(BLUPs.GY=coef(model1)$Genotype[,1]) #Genotype values (Blups +Intercept)
```

Heritability

- Here will show how to calculate the heritability. Two approaches will be show how to estimate heritability: 1) Based on Variance components and 2) Based on Cullis et al. 2006 is also ….$1 - \frac{\overline{V}_{BLUp}}{2\sigma^2 g}$. Where $\overline{V}_{BLUP}$ is mean variance difference of two genotypes based on BLUPs and $\sigma^2 g$ is variance of genotypes.

Heritability Based on Varaince Components

```
> # Extract the variance components
>   Ve<- data.frame (VarCorr(model1))
>   Ve
```

grp	var1	var2	vcov	sdcor
Genotype	(Intercept)	NA	431860.94	657.1613
Rep:Block	(Intercept)	NA	28498.55	168.8151
Residual	NA	NA	193254.94	439.6077

```
> # Now calculate heritability using variance components
>   genotype.var=Ve[1,4]
>   error.var=Ve[2,4]
> # Now heritability
>   h2=genotype.var/(genotype.var+error.var)*100
>   h2
```

```
[1] 93.8095
```

Relaibility based on Cullis et al.

```
> # Reliability
>   std.err<-se.ranef(model1)$Genotype
>   v_BLUP<- mean(std.err)
> # Heritability/Reliability
>   h2<- (1-((v_BLUP)^2/(Ve[1,4]*2)))*100
>   h2
```

```
[1] 90.55036
```

Heritability Based on MET Analysis

Model 2.lme4

- Here we will analyze all the environments jointly and extract the single BLUPs for each genotype.
- We will use mixed model analysis in lme4 r package model.
- We will treat genotypes as random and environment as fixed effect.

Combined ANOVA

- Here ANOVA will be generated for all the factor levels.
- Replications are nested with environments and Blocks are within Replications which are nested within environment.

```
> # Linear model to get ANOVA
>   demo.data.filtered$Environment<-as.factor(demo.data.filtered$Environment)
>   model.anova<-lm(formula = Yield~Genotype+Environment+Genotype*Environment+Environment:Rep+ Environme
+ data=demo.data.filtered)
> # Get ANOVA
>   anova(model.anova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	199	547177221	2749634	5.577295	0.00e+00
Environment	3	2081940510	693980170	1407.653397	0.00e+00
Genotype:Environment	588	1002589262	1705084	3.458553	0.00e+00
Environment:Rep	4	22794162	5698541	11.558788	0.00e+00
Environment:Rep:Block	32	37568681	1174021	2.381358	3.54e-05
Residuals	754	371725774	493005	NA	NA

Significant differences are observed for all factors and genotype by environment interactions are significant

Combined Analysis in lme4

- The model we will use is give below:

$y_{ijkl} = \mu + g_i + e_j + (ge)_{ij} + r_{jk} + b_{jkl} + \epsilon_{ijklm}$

$\mu =$ overall mean

$g_i =$ random effect of the ith genotype

$e_j =$ random effect of the jth environment

$(ge)_{ij} =$ is the interaction effect of ith genotypes with the jth environment

$r_{jk} =$ fixed effect of the kth replication nested within jth environment

$b_{jkl} =$ random effect of lth block nested with j environment and kth replication

$\varepsilon_{ijkl} =$ residual error

here we assume residuals are independent and identically distributed

- Mixed models are powerful tools to handle assumptions of linear model Read this one
- We will extract variance components and also calculate heritability.

```
> demo.data.filtered$Environment<-as.factor(demo.data.filtered$Environment)
> Model3.lme4<-lmer(Yield~Rep+(1|Genotype)+(1|Environment)+
+           (1|Block), data=demo.data.filtered)
```

Summary of MET results

- In summary we will get following summarized results: 1) Description of model we used, 2) Random effects and varainces, 3) Fixed effects, 4) Correlation of fixed effects

```
>   summary(Model3.lme4)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: Yield ~ Rep + (1 | Genotype) + (1 | Environment) + (1 | Block)
   Data: demo.data.filtered

REML criterion at convergence: 26571.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7537 -0.5702 -0.0115  0.4669  5.3187

Random effects:
 Groups      Name        Variance Std.Dev.
 Genotype    (Intercept)  205131   452.9
 Block       (Intercept)    7224    85.0
 Environment (Intercept) 1789049  1337.6
 Residual                1033082  1016.4
Number of obs: 1581, groups:  Genotype, 200; Block, 5; Environment, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)  4740.65     671.60   7.059
```

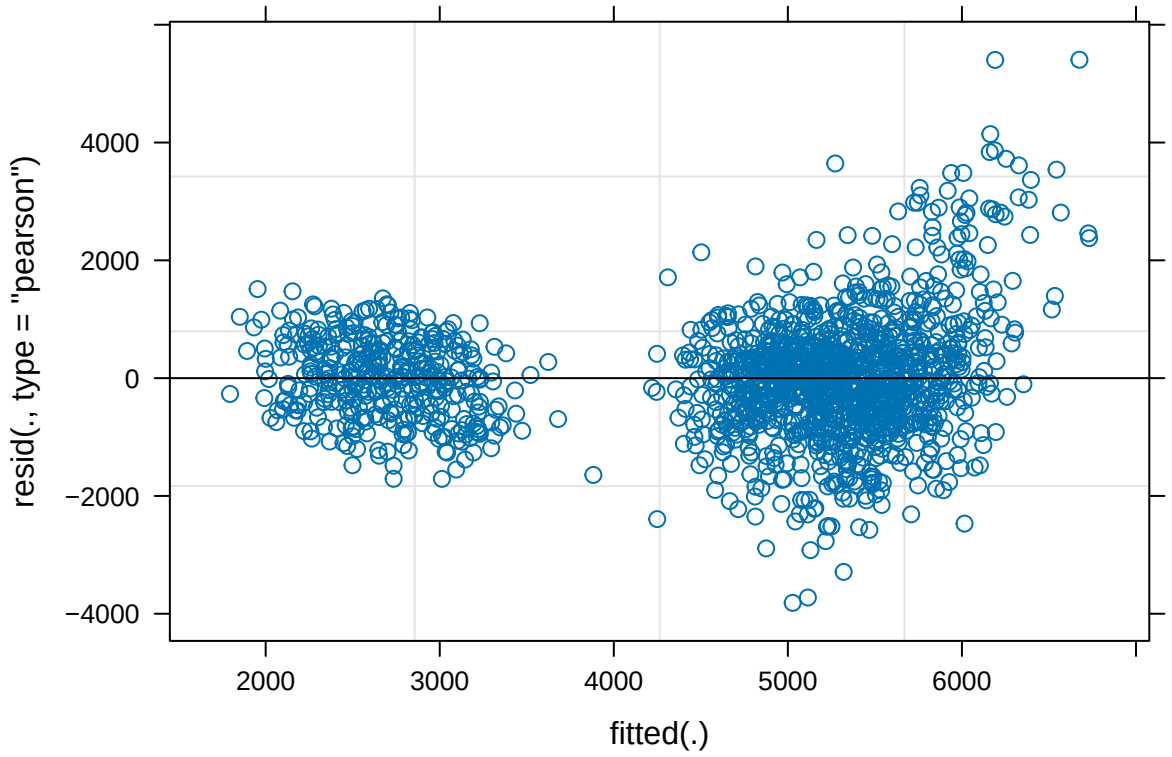
```
Rep2         -151.88      51.13  -2.971

Correlation of Fixed Effects:
     (Intr)
Rep2 -0.038
```

Plot of model

- With the plot function model we will get the residuals vs fitted values

```
>    plot(Model3.lme4)
```



Relaibility Using Mean Difference

- Based on **Cullis et al.2006**

```
> # Extract the variance components
>    Ve<- data.frame (VarCorr(Model3.lme4))
>    Ve
```

grp	var1	var2	vcov	sdcor
Genotype	(Intercept)	NA	205131.050	452.91395
Block	(Intercept)	NA	7224.397	84.99645
Environment	(Intercept)	NA	1789048.808	1337.55329
Residual	NA	NA	1033082.471	1016.40665

```
>    std.err<-se.ranef(Model3.lme4)$Genotype
>    v_BLUP<- mean(std.err)
> # Heritability/Reliability
```

```
>   h2<- (1-((v_BLUP)^2/(Ve[1,4]*2)))*100
>   h2
```

```
[1] 80.29987
```

Additional Resources

- Here In this section we have provided additional R resources where similar analysis and mixed models can be run:
- Fitting linear mixed-effects models using lme4
- Linear Mixed-Effects Models Using R
- sommer: Solving Mixed Model Equations in R
- robustlmm
- Introduction to Linear Mixed Models
- Computing Heritability and Selection Response From Unbalanced Plant Breeding Trials
- Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials

---

For any suggestions or comments, please feel to reach at waseem.hussain@cgiar.org; m.anumalla@cgiar.org; m.catolos@cgiar.org

---

If your experiment needs a statistician, you need a better experiment - Ernest Rutherford