Introduction to Genomic Selection in R using the rrBLUP Package

Amy Jacobson jaco0795@umn.edu University of Minnesota



Learning Objectives

- Download the package and load the sample files
- Impute missing markers using A.mat()
- Define the training and validation populations
- Run mixed.solve() and determine accuracy of predictions

Overview of rrBLUP package

- Download from CRAN-version 4
 - Must use R version 2.14.1 or greater
- Uses ridge regression BLUP for genomic predictions
- Predicts marker effects through mixed.solve()
- A.mat() command can be used to impute missing markers
 - Mixed.sove does not allow NA marker values
- Define the training and validation populations

One Step vs. Two Step

One step

- Uses a mixed model analysis for the plot data
- Two step
 - Adjusted means are calculated across locations
 - Means are then used in ridge regression blup
- This webinar uses a two step approach
 - Computationally more efficient and faster

Install the rrBLUP Package Launch R->Packages->Install Package

Select CRAN Mirror nearest you

🥂 RGui (64-bit)	-		And shares have									
File Edit View Misc Pac	kages Windows Help											
🖻 🖪 🖪 🛍 🕻												
R Console	Set CRAN mirror Select repositories											
	Install package(s)		^									
R version 2.15.1 Copyright (C) 201	Update packages	omputing										
ISBN 3-900051-07- Platform: x86 64-	Install package(s) from local zip files											
You are welcome to r Type 'license()' or Natural language s R is a collaborative Type 'contributors() 'citation()' on how Type 'demo()' for so 'help.start()' for a Type 'q()' to quit F [Previously saved wo		conditions. details. ish locale rs. plications. = help, or										
>			-									

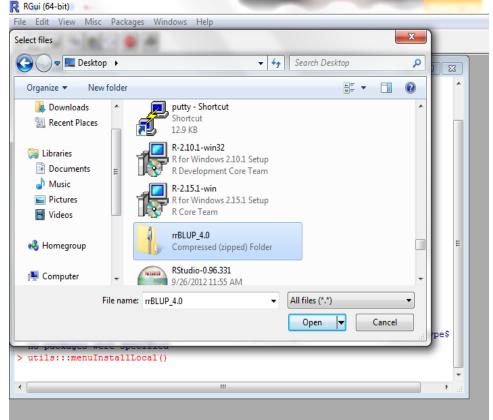
Select the rrBLUP package

R Console	
-	Packages
version 2.15.1 (2012-06-22) "Roasted Marshmallows"	RODA
opyright (C) 2012 The R Foundation for Statistical Computing	rqmcmb2
SBN 3-900051-07-0	Rquake
latform: x86_64-pc-mingw32/x64 (64-bit)	RQuantLib
	Rramas rrBLUP
is free software and comes with ABSOLUTELY NO WARRANTY.	rrBlupMethod6
ou are welcome to redistribute it under certain conditions.	rrcov
<pre>'ype 'license()' or 'licence()' for distribution details.</pre>	rrcovHD
Natural language support but running in an English locale	rrcovNA
Natural language support but funning in an English locale	rrdf
is a collaborative project with many contributors.	rrdflibs
ype 'contributors()' for more information and	rredis rreval
citation()' on how to cite R or R packages in publications.	RRF
	rriskBayes
<pre>'ype 'demo()' for some demos, 'help()' for on-line help, or</pre>	rriskDistributions
help.start()' for an HTML browser interface to help.	rrlda
ype 'q()' to quit R.	rrules
	rsae
Previously saved workspace restored]	RSAGA
	rscproxy rsdepth
utils:::menuInstallPkgs()	RSearchYJ
Please select a CRAN mirror for use in this session	RSeed
	rseedcalc
	RSEIS
	rsem
	Rserve rSFA
	rsqcc
	Rsge
	OK Cancel

- Install the package by a zip file
- <u>http://cran.r-</u> project.org/web/packa ges/rrBLUP/index.html
- Packages->install package from local zip files

<pre>ile Edit View Misc Packages Windows Help Cody packages Windows Help Copyright (C) 20 ISBN 3-900051-07 Platform: x86_64 R is free softwal Ntail package(s) Update packages Yup date packages Nutural language support but running in an English locale R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications. Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. [Previously saved workspace restored] > utils:::menuInstallPkgs() Please select a CRAN mirror for use in this session Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type\$ no packages were specified > </pre>	RGui (64-bit)	Concentration and		
<pre>k R Console Copyright (C) 20 Install package(s) Update package(s) Update package(s) Update package(s) Update package(s) from local zip files NTY. You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details. Natural language support but running in an English locale R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications. Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. [Previously saved workspace restored] > utils:::menuInstallPkgs() Please select a CRAN mirror for use in this session Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type\$ no packages were specified > </pre>	ile Edit View Misc Pac	kages Windows Help		
R Console Copyright (C) 201 ISBN 3-900051-07- Platform: x86_64 R is free softwat Install package(s) from local zip files NTY. You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details. Natural language support but running in an English locale R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications. Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. [Previously saved workspace restored] > utils:::menuInstallPkgs() Please select a CRAN mirror for use in this session Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type\$ no packages were specified	é 🖞 🖬 🛍 🛍 🕻	Load package		
Copyright (C) 20 ISBN 3-900051-07 Platform: x86_64 R is free softwa You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details. Natural language support but running in an English locale R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications. Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. [Previously saved workspace restored] > utils:::menuInstallPkgs() Please select a CRAN mirror for use in this session Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type\$ no packages were specified >	R Console		_ 0	×
<pre>x Is Tree Soltwal</pre>	ISBN 3-900051-07-	Install package(s)	Computing	
<pre>R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications. Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. [Previously saved workspace restored] > utils:::menuInstallPkgs() Please select a CRAN mirror for use in this session Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type\$ no packages were specified > </pre>	You are welcome to :	redistribute it under certain	conditions.	
<pre>Figure Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. [Previously saved workspace restored] > utils:::menuInstallPkgs() Please select a CRAN mirror for use in this session Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type\$ no packages were specified ></pre>	R is a collaborative Type 'contributors()	e project with many contributo)' for more information and	rs.	
<pre>> utils:::menuInstallPkgs() Please select a CRAN mirror for use in this session Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type\$ no packages were specified > </pre>	'help.start()' for a	an HTML browser interface to h		Ш
Please select a CRAN mirror for use in this session Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type\$ no packages were specified	[Previously saved wo	orkspace restored]		
· · · · · · · · · · · · · · · · · · ·	Please select a Error in install.pac no packages were s	CRAN mirror for use in this so ckages(NULL, .libPaths()[1L],		eŞ
	- 1			

 Select the package from saved location



 Now that the package is installed, the library must be loaded every time R is opened

I	RGui (64-bit)	
	File Edit View Misc Packages Windows Help	
ĺ	R Console	
	<pre>'citation()' on how to cite R or R packages in publications.</pre>	
	Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.	
	[Previously saved workspace restored]	
	> #load in the sample files > library(rrBLUP)	

Sample Files

- Files downloaded from the Hordeum Toolbox <u>http://hordeumtoolbox.org/</u>
- University of Minnesota barley breeding program preliminary yield trail-St. Paul location in 2009
- Phenotypic traits-yield, plant height and heading date
- 1178 markers, 164 NA markers
- 1 = homozygous for parent 1, 0 = heterozygous, and -1 homozygous for parent 2
 - Markers must be in the {-1,0,1} format for rrBLUP

- Setwd()-Set the working directory to the location of the sample files
- Read.table command used for .txt files
- Read.csv command used for .csv files
- Header=F since sample marker file does not have a header with marker names

RGui (64-bit)	
File Edit View Misc Packages Windows Help	
R Console	
<pre>> setwd("C://Users//Amy J//Documents//webinar") > Markers <- as.matrix(read.table(file="snp.txt"), header=FALSE)</pre>	*

- head() command used to see the first 5 lines of a file
- Useful to see if data was loaded correctly

	RGui (64-bit) File Edit View Misc Packages Windows Help																				
R R C	🥂 R Console															x					
> he	ad (M	arke	ers)																		1
				4 V	5 1	76 V.	7 V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	
[1,]	-1	1	1 -	1 -	1	1 -1	1 1	1	1	-1	1	1	1	-1	-1	-1	1	-1	1	1	
[2,]	-1	1	1 -	1 -	1	1 -1	L 1	1	1	1	1	1	1	1	-1	-1	1	-1	1	1	
[3,]	-1	1	1 -	1 -	1	1 -1	L -1	-1	1	-1	1	1	1	-1	-1	-1	1	-1	1	1	
[4,]	-1	1	1 -	1 -	1	1 1	L 1	1	1	-1	1	1	1	-1	-1	-1	1	-1	1	1	
[5,]	1	1	1	1 -	1	1 -1	l -1	-1	1	1	1	1	1	1	-1	-1	1	-1	1	1	
[6,]	-1	1	1 -	1 -	1	1 -1	L -1	-1	1	-1	1	1	1	-1	-1	-1	1	-1	1	1	
	V22	V23	3 V2	4 V	25	V26	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36	V37	V38	V39		
[1,]	-1	1	L –	1	1	-1	-1	-1	. 1	1	-1	1	. 1	. 1	-1	1	-1	. 1	1		
[2,]	-1	1	L –	1	1	-1	-1	1	. 1	1	-1	1	-1	. 1	-1	-1	-1	. 1	1		
[3,]	-1	1	L -	1	1	1	-1	-1			-1	1	-1	1	-1	-1	-1	. 1	1		
[4,]	-1		_	1	1	-1	-1	1		_					_	1					
[5,]	-1	1	-	1	1	1	-1	1		_	_	-	-	-	_	1			_		
[6,]	-1	1		1	1	-1	-1	1	. 1	1	-1	1	. 1	. 1	-1	-1	-1	-	_		-
	V40	V41	L V4	2 V	43	V44	V45	V46	V47	V48	V49	V50	V51	V52	V53	V54	V55	V56			
[1,]	-1	1	L	1	-1	-1	-1	-1	. 1	1	1	. 1	-1	1	-1	-1	1	. 1	1		
[2,]	-1	1	L	1	-1	-1	-1	-1	. 1	1	1	. 1	-1	1	-1	-1	-1	. 1	1		
*																					
<u> </u>				_	_																

- Load the phenotype file and use the head command to see the first five lines
 - Header=T since phenotype files have column names
- Markers and phenotypes must be in matrix format

R RC	G ui (64	-bit)	-					_		-	-	-	Ξ.	
File	Edit	View	Misc	Packages	Windows	Help								
2	c I			€ 👜 ב	3									
R	R Cons	ole												
[6,	-	_	1	-1	-1	1	-1	1	-1	0	1	1	1	^
[1,		176 V -1	-1											
	i		1											
		-1												F
		-1												
[5,	1	-1	-1											:
[6,	1	-1	-1											•
> I	heno	<-as	.matr	ix(read	.table(:	file =	="trai	its.txt	", hea	der=TR	JE))			
> h	nead (Pheno	o)											1
	gr	ain_y	yield	pht_hei	ght Head	ding_l	Date							1
[1,	1		4869		78		57.5							•
[2,	1		4621		74	;	56.0							:
[3,	1		4557		66	;	58.0							•
[4,	1		5484		77	;	59.0							
[5,	1		4641		69	;	55.5							
[6,	1		4516		76	;	58.5							
>														
•														▶ at

- Determine the size of the matrices
- dim() command gives the number of rows and columns
- 96 observations and 1178 markers, 3 traits

```
RGui (64-bit)
File Edit View Misc Packages Windows Help
 🛎 🖆 🕞 🖻 🖀 🚭 🥌
 R Console
               -1 -1
 [5,]
          1
               -1
 [6,]
          1
                     -1
 5
 > Pheno <-as.matrix(read.table(file ="traits.txt", header=TRUE))
 > head(Pheno)
      grain yield pht height Heading Date
             4869
                          78
                                      57.5
 [1,]
 [2,]
             4621
                          74
                                      56.0
 [3,1
             4557
                          66
                                      58.0
             5484
                         77
                                      59.0
 [4, ]
 [5,1
             4641
                          69
                                      55.5
                        76
                                      58.5
 [6,1
             4516
   dim(Markers)
       96 1178
 > dim(Pheno)
 [1] 96 3
 >
```

Learning Objectives

- Download the package and load the sample files
- Impute missing markers using A.mat()
- Define the training and validation populations
- Run mixed.solve() and determine accuracy of predictions

- rrBLUP mixed.solve() does not allow for missing markers
- Imputed value is the population mean for that marker
- Useful for SNP data since level of missing data is low
 - In the sample files 164 markers are missing out of 1178 (0.14%)
- A.mat also calculates the additive relationship matrix

- max.missing-maximum proportion of missing data
 - If 50% of markers are missing data then markers are not imputed
- impute method- imputes the mean of the markers
- return.imputed-prints out the imputed results if set to TRUE

```
> #what if markers are NA?
> #impute with A.mat
> impute=A.mat(Markers,max.missing=0.5,impute.method="mean",return.imputed=T)
> |
```

- >impute=A.mat(Markers,max.missing=0.5,imput e.method="mean",return.imputed=T)
- > Markers_impute=impute\$imputed
 - Rename imputed marker matrix as Markers_impute
- impute\$imputed-returns the imputed marker matrix
- impute\$A-returns the additive relationship matrix

>impute\$imputed

Imputed marker value

	<u>R</u> RGui	(64-bit))				-			-		-					
	File E	dit Vie	w Mi	sc Pac	:kages	Windo	ws H	elp									
	é		B) (*	1	•	3											
	🖳 R C	onsole										4					
		V137	V138	V139	V140	V141	V142	V143	V144	V145		V146	V147	V148	V149	V150	*
	[1,]	-1	1	1	1	1	-1	-1	1		0.80		1 1	1	1	1	
	[2,]	-1	1	1	1	1	-1	-1	-1		1.00		1	1	1	1	
	[3,]	-1	1	1	1	1	-1	-1	1	-1	1.00	00000	1	1	1	1	
Marker	[4,]	1	-1	1	1	1	-1	-1	1		1.00		1	1	1	1	
voluo	[5,]	-1	1	1	1	1	-1	-1	-1		1.00		1	-1	1	1	
value	[6,]	-1 V151	1 V152	V153	V154	1 V155	-1 V156	-1 V157	-1 V159	-1 V150	1.00 V160	V161	1 V162	1 V163	1 V164	V165	
left NA if	[1,]	1	1 1	-1	1	-1	1	1	-1	1	1 1	-1	-1	1	1	-1	
moro	[2,]	1	1	-1	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	
more	[3,]	1	1	-1	1	-1	1	1	-1	1	1	-1	-1	1	1	-1	
than	[4,]	1	1	-1	1	-1	1	1	-1	1	1	-1	-1	1	1	-1	
	[5,]	1	1	-1 -1	1	-1 -1	1	1	-1 -1	1	1	-1	-1	1	1	1	
50%	[6,]	V166	V167	V168	v169	V170	V171	V172	V173	V174	V175	-1 V176	-1 V177	V178	¥179	V180	
missing	[1,]	1	1	.100	NA	1	1	1	-1	-1	1	1	1	-1	-1	1 1	
•	[2,]	1	1	1	NA	1	1	1	-1	-1	1	1	1	-1	-1	1	
data 🗕	[3,]	1	1	→1	NA	1	1	1	-1	1	1	1	1	-1	-1	1	
	[4,]	1	1	1	NA	1	1	1	-1	-1	1	1	1	-1	-1	1	-

- Remove markers that had more than 50% missing data
 - NA values are not allowed in mixed.solve
 - Two markers in the SNP file must be removed
 - Column 169 and 562
 - New dimensions show 2 less columns
- Use Markers_impute2 as marker matrix for estimating marker effects

```
> Markers_impute2=Markers_impute[,-c(169,562)]
> dim(Markers_impute)
[1] 96 1178
> dim(Markers_impute2)
[1] 96 1176
> |
```

Learning Objectives

- Download the package and load the sample files
- Impute missing markers using A.mat()
- Define the training and validation populations
- Run mixed.solve() and determine accuracy of predictions

- Training population-genotyped and phenotyped
- Validation population-phenotype values estimated based on marker effects calculated from training population
- Code is set that 60% of the total population is the training population
 - 40% validation population

- 58 (60% of total population of 96) random numbers sampled to determine which individuals are in the training population
- Individuals are the row numbers for the phenotypes and marker matrices
- Sampled numbers will be different every time the code is run and will affect the correlation accuracy

```
> train= as.matrix(sample(1:96, 58))
> head(train)
      [,1]
[1,] 52
[2,] 82
[3,] 50
[4,] 14
[5,] 7
[6,] 80
> |
```

- Validation population is 40% of the total population
- setdiff() command determines the numbers that are not in the training population and will be part of the validation population

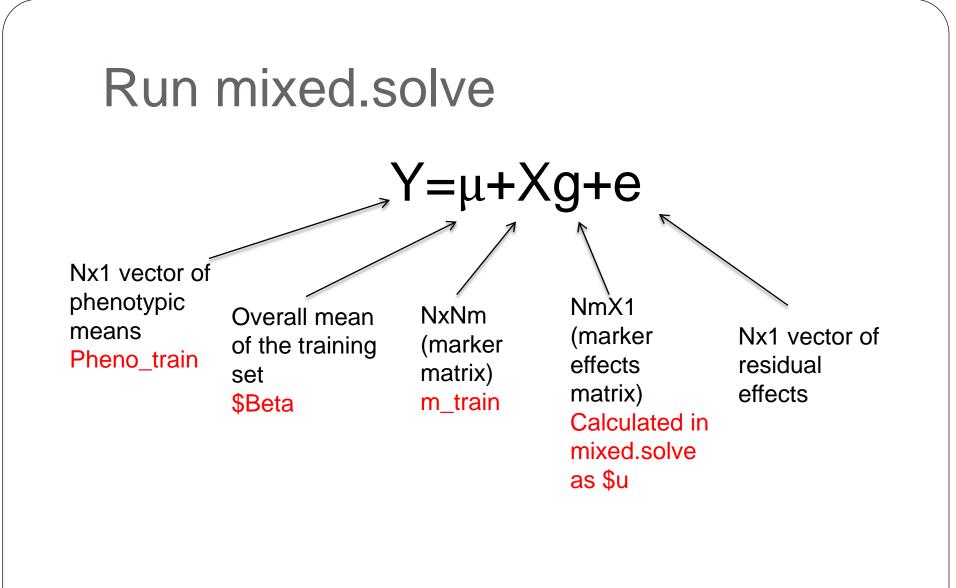
```
> test<-setdiff(1:96,train)
> test
[1] 6 12 18 19 22 23 26 27 28 29 33 34 36 40 41 43 47 48 53 54 55 56 57 58 59 62 66
[28] 68 71 75 77 79 83 84 86 90 91 96
>
```

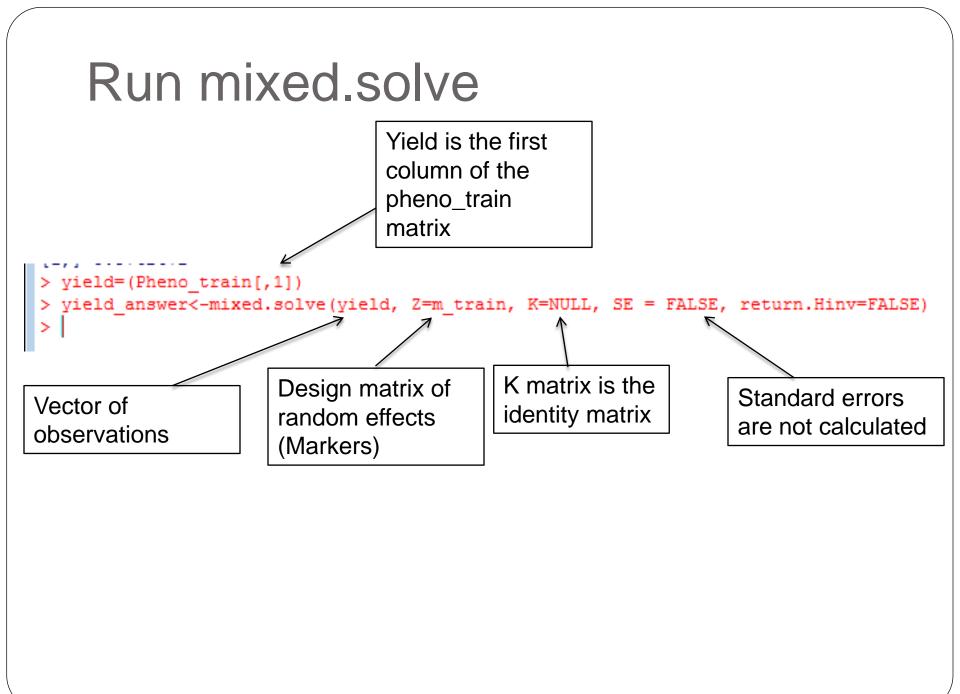
- Pheno_train and m_train are the phenotype and marker matrices for the values in the training population
- Pheno_valid and m_valid will be the validation populations

```
> Pheno_train=Pheno[train,]
> m_train=Markers_impute2[train,]
> Pheno_valid=Pheno[test,]
> m_valid=Markers_impute2[test,]
> |
```

Learning Objectives

- Download the package and load the sample files
- Impute missing markers using A.mat()
- Define the training and validation populations
- Run mixed.solve() and determine accuracy of predictions





Run mixed.solve

- Yield_answer\$u is the output of the marker effects
- head(e) shows the marker effects for the first five markers

```
> yield_answer<-mixed.solve(yield, Z=m_train, K=NULL, SE = FALSE, return.Hinv=FALSE)
> YLD = yield_answer$u
> e = as.matrix(YLD)
> head(e)
        [,1]
V1 1.1380597
V2 -0.1141220
V3 0.4970927
V4 2.2986051
V5 0.3579770
V6 -0.1141220
>
```

Run mixed.solve

- m_valid*e = marker validation matrix times the marker effects
- Pred_yield=predicted yield based on the marker effects of the training population with the grand mean added in

```
> pred_yield_valid = m_valid %*% e
> pred_yield=(pred_yield_valid[,1])+yield_answer$beta
> pred_yield
[1] 4745.698 4621.133 4742.935 4601.210 4671.582 4636.899 4552.350 4486.954
[9] 4589.440 4601.534 4508.288 4656.675 4462.313 4493.898 4668.741 4498.701
[17] 4708.654 4593.296 4441.527 4705.500 4597.538 4089.056 4177.749 4261.560
[25] 4107.757 4207.431 4454.215 4713.850 4740.123 4537.690 4585.838 4526.935
[33] 4570.133 4512.558 4613.167 4412.658 4747.170 4872.127 4774.157 4697.992
[41] 4640.538 4576.519 4707.957 4658.228 4772.145 4596.747 4371.145 4779.256
[49] 4427.464 4525.557 4305.716 4564.654 4450.188 4634.591 3989.726 4068.685
[57] 4043.495 3886.869
```

- Correlation between the predicted yield values and the observed yield values
- Accuracy will change slightly each time due to different individuals sampled for the training and validation populations

```
> yield_valid = Pheno_valid[,1]
> YLD_accuracy <-cor(pred_yield_valid, yield_valid, use="complete")
> YLD_accuracy
       [,1]
[1,] 0.2521498
> |
```

Plant Height

```
> PHT_HT=(Pheno_train[,2])
> PHT_HT_answer<-mixed.solve(PHT_HT, Z=m_train, K=NULL, SE = FALSE, return.Hinv=FALSE)
> PHT_HT = PHT_HT_answer$u
> e = as.matrix(PHT_HT)
> pred_PHT_HT_valid = m_valid %*% e
> PHT_HT_valid = Pheno_valid[,2]
> PHT_HT_accuracy <-cor(pred_PHT_HT_valid, PHT_HT_valid, use="complete")
> PHT_HT_accuracy
        [,1]
[1,] 0.4055428
> |
```

Heading Date

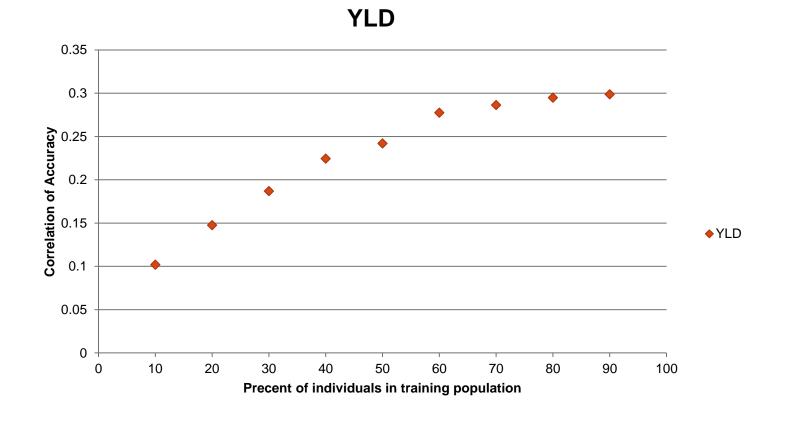
```
> HD_DATE=(Pheno_train[,3])
> HD_DATE_answer<-mixed.solve(HD_DATE, Z=m_train, SE = FALSE, return.Hinv=FALSE)
> HD_DATE = HD_DATE_answer$u
> e = as.matrix(HD_DATE)
> pred_HD_DATE_valid = m_valid %*% e
> HD_DATE_valid = Pheno_valid[,3]
> HD_DATE_accuracy <-cor(pred_HD_DATE_valid, HD_DATE_valid, use="complete")
> HD_DATE_accuracy
        [,1]
[1,] 0.5205029
> |
```

Correlation accuracy with 500 iterations

```
> #### cross validation for many cycles for yield only
> traits=1
> cycles=500
> accuracy = matrix(nrow=cycles, ncol=traits)
> for(r in 1:cycles)
+ {
+ train= as.matrix(sample(1:96, 38))
+ test<-setdiff(1:96,train)
+ Pheno train=Pheno[train,]
+ m train=Markers impute2[train,]
+ Pheno valid=Pheno[test,]
+ m valid=Markers impute2[test,]
+ yield=(Pheno train[,1])
+ yield answer<-mixed.solve(yield, Z=m train, K=NULL, SE = FALSE, return.Hinv=F$
+ YLD = yield answer$u
+ e = as.matrix(YLD)
+ pred yield valid = m valid %*% e
+ pred yield=(pred yield valid[,1])+yield answer$beta
+ pred yield
+ yield valid = Pheno valid[,1]
+ accuracy[r,1] <-cor(pred yield valid, yield valid, use="complete" )
+ }
> mean(accuracy)
[1] 0.2305713
```

- Correlation accuracy is different for each trait
- Values will be different every time it is run since different lines will be included in the training or validation sets
- Accuracy is affected by training size, validation size, number of markers and heritability

Effects of training population size on accuracy



```
    Headers incorrectly input
```

```
> Pheno <-as.matrix(read.table(file ="traits.txt", header=F))
> head(Pheno)
     V1
                   V2
                                 V3
[1,] "grain yield" "pht height" "Heading Date"
[2,] "4869"
                   "78"
                                 "57.5"
[3,1
     "4621"
                  "74"
                                 "56"
[4.1
     "4557"
                 "66"
                                 "58"
                  "77"
[5,] "5484"
                                 "59"
                   "69"
                                "55.5"
[6,] "4641"
> train= as.matrix(sample(1:96, 38))
> test<-setdiff(1:96,train)</pre>
> Pheno train=Pheno[train,]
> m train=Markers impute2[train,]
> Pheno valid=Pheno[test,]
> m valid=Markers impute2[test,]
> yield=(Pheno train[,1])
> yield answer<-mixed.solve(yield, Z=m train, K=NULL, SE = FALSE, return.Hinv=F$
Error in crossprod(x, y) :
  requires numeric/complex matrix/vector arguments
```

NA Markers

```
> train= as.matrix(sample(1:96, 38))
> test<-setdiff(1:96,train)
> Pheno_train=Pheno[train,]
> m_train=Markers[train,]
> Pheno_valid=Pheno[test,]
> m_valid=Markers[test,]
> yield=(Pheno_train[,1])
> yield_answer<-mixed.solve(yield, Z=m_train, K=NULL, SE = FALSE, return.Hinv=FALSE)
Error in eigen(Hb, symmetric = TRUE) : infinite or missing values in 'x'</pre>
```

Incorrect matrix dimensions

Removed one individual from phenotype matrix

```
> ########
> #define the training and test populations
> #training-60% validation-40%
> train= as.matrix(sample(1:96, 38))
> test<-setdiff(1:96,train)
> Pheno train=Pheno[train,]
> m train=Markers impute2[train,]
> Pheno valid=Pheno[test,]
Error: subscript out of bounds
> m valid=Markers impute2[test,]
> ########
> yield=(Pheno train[,1])
> yield answer<-mixed.solve(yield, Z=m train, K=NULL, SE = FALSE, return.Hinv=F$
> YLD = yield answer$u
> e = as.matrix(YLD)
> pred yield valid = m valid %*% e
> pred yield=(pred yield valid[,1])+yield answer$beta
> yield valid = Pheno valid[,1]
Error: object 'Pheno valid' not found
> YLD accuracy <-cor(pred yield valid, yield valid, use="complete" )
Error in is.data.frame(y) : object 'yield valid' not found
> YLD accuracy
Error: object 'YLD accuracy' not found
```

- Read in values as characters instead of numeric
 - Quotes around values

```
> Pheno <-as.matrix(read.table(file ="traits.txt", header=F))</p>
> head(Pheno)
                   \mathbf{V2}
                                 V3
     V1
     "grain yield" "pht height" "Heading Date"
                  "78"
                                 "57.5"
     "4869"
                 "74"
                                 "56"
                 "66"
                                 "58"
                  "77"
     #5484#
                                 #59#
```

Resources

- rrBLUP reference manual
 - <u>http://cran.r-</u> project.org/web/packages/rrBLUP/rrBLUP.pdf
- rrBLUP vingettes
 - <u>http://cran.r-</u> project.org/web/packages/rrBLUP/vignettes/vignette
 <u>.pdf</u>
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250-255. doi: 10.3835/plantgenome2011.08.0024

Acknowledgements

Rex Bernardo Emily Combs Lian Lian Chris Schaefer Lisa-Marie Krchov

rrBLUP Jeff Endelman

Funding Monsanto Company USDA SolCAP



United States Department of Agriculture National Institute of Food and Agriculture SolCAP David Francis Shawn Yarnes John McQueen

Dataset TCAP Hordeum's Toolbox

ΜΟΝ SΑΝΤΟ



Questions?