Using asremlPlus, in conjunction with asreml, to do a linear mixed model analysis of a wheat experiment using hypothesis tests

Chris Brien

06 January, 2025

This vignette shows how to use asremlPlus (Brien, 2025), in conjunction with asreml (Butler et al., 2020), to employ hypothesis tests to select the terms to be included in a mixed model for an experiment that involves spatial variation. It also illustrates diagnostic checking and prediction production and presentation for this experiment. Here, asremlPlus and asreml are packages for the R Statistical Computing environment (R Core Team, 2025).

It is divided into the following main sections:

- 1. Set up the maximal model for this experiment
- 2. Perform a series of hypothesis tests to select a linear mixed model for the data
- 3. Diagnostic checking using residual plots and variofaces
- 4. Prediction production and presentation

1. Set up the maximal model for this experiment

```
library(knitr)
opts_chunk$set("tidy" = FALSE, comment = NA)
suppressMessages(library(asreml, quietly=TRUE))

## Offline License checked out Mon Jan 6 22:26:44 2025

packageVersion("asreml")

## [1] '4.2.0.332'
suppressMessages(library(asremlPlus))
packageVersion("asremlPlus")

## [1] '4.4.44'
suppressMessages(library(qqplotr, quietly=TRUE))
options(width = 100)
```

Get data available in asremlPlus

The data are from a 1976 spring wheat experiment and are taken from Gilmour et al. (1995). An analysis is presented in the asrem1 manual by Butler et al. (2020, Section 7.6), although they suggest that it is a barley experiment.

```
data(Wheat.dat)
```

Fit the maximal model

In the following a model is fitted that has the terms that would be included for a balanced lattice. In addition, a term WithinColPairs has been included to allow for extraneous variation arising between pairs of adjacent lanes. Also, separable ar1 residual autocorrelation has been included. This model represents the maximal anticipated model,

```
ASReml Version 4.2 06/01/2025 22:26:44
```

	LogLik	Sigma2	DF	wall		
1	-724.1213	23034.14	124	22:26:44		
2	-717.4149	9206.931	124	22:26:44	(2 restrained)
3	-694.8752	26492.99	124	22:26:44	(2 restrained)
4	-694.1600	33101.80	124	22:26:44	(1 restrained)
5	-692.0020	36912.26	124	22:26:44	(1 restrained)
6	-691.7892	46701.51	124	22:26:44	(2 restrained)
7	-691.8336	46208.51	124	22:26:44	(1 restrained)
8	-691.7749	47698.26	124	22:26:44		
9	-691.7711	47041.85	124	22:26:44		

Warning in asreml(yield \sim WithinColPairs + Variety, random = \sim Rep/(Row + : Some components changed by more than 1% on the last iteration

The warning from asreml is probably due to a bound term.

Initialize a testing sequence by loading the current fit into an asrtests object

A label and the information criteria based on the full likelihood (Verbyla, 2019) are included in the test.summary stored in the asrtests object.

```
Warning in infoCriteria.asreml(asreml.obj, IClikelihood = ic.lik, bound.exclusions = bound.exclusions):
Rep
```

Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Log-likelihood not converged

Check for and remove any boundary terms

```
current.asrt <- rmboundary(current.asrt, IClikelihood = "full")</pre>
Warning in infoCriteria.asreml(asreml.obj, IClikelihood = ic.lik): The following bound terms were disco
ASReml Version 4.2 06/01/2025 22:26:45
          LogLik
                       Sigma2
                                  DF
                                          wall
                      47071.42
 1
       -691.7710
                                  124
                                        22:26:45
Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Log-likelihood
not converged
summary(current.asrt$asreml.obj)$varcomp
                                                   z.ratio bound %ch
                         component
                                      std.error
Rep:Row
                      4.293282e+03 3.199458e+03 1.3418779
                                                               P 0.0
Rep:Column
                      1.575689e+02 1.480357e+03 0.1064398
                                                               P 0.7
units
                      5.742689e+03 1.652457e+03 3.4752438
                                                               P 0.0
                      4.706787e+04 2.515832e+04 1.8708669
Row:Column!R
                                                               P 0.0
Row:Column!Row!cor
                      7.920301e-01 1.014691e-01 7.8056280
                                                               U 0.0
Row:Column!Column!cor 8.799559e-01 7.370402e-02 11.9390486
                                                               U 0.0
print(current.asrt, which = "testsummary")
```

Sequence of model investigations for yield

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

```
terms DF denDF p AIC BIC action 1 Maximal model 26 6 NA 1646.129 1742.47 Starting model 2 Rep 1 NA NA 1646.129 1742.47 Boundary
```

Rep has been removed because it has been constrained to zero. Following the recommendation of Littel et al. (2006, p. 150), the bound on all variance components is set to unconstrained (U) using setvariances.asreml so as to avoid bias in the estimate of the residual variance. Alternatively, one could move Rep to the fixed model.

Unbind Rep, Row and Column components and reload into an asrtests object

```
DF
         LogLik
                       Sigma2
                                         wall
                                       22:26:46
1
      -724.1213
                     23034.14
                                 124
 2
      -717.4149
                                       22:26:46 ( 2 restrained)
                     9206.931
                                 124
 3
      -694.8752
                     26492.99
                                 124
                                       22:26:46
                                                 ( 2 restrained)
 4
      -693.9744
                     33129.65
                                 124
                                       22:26:46 ( 1 restrained)
 5
      -692.8856
                     39662.12 124
                                       22:26:46
 6
      -691.4276
                                 124
                                       22:26:46
                     53103.83
 7
      -691.2387
                     48092.17
                                 124
                                       22:26:46
                                       22:26:46
 8
                     47278.94
      -691.1808
                                 124
9
      -691.1710
                     46850.98
                                 124
                                       22:26:46
                                       22:26:46
10
      -691.1700
                     46690.46
                                 124
Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Some components
changed by more than 1% on the last iteration
WARN [2025-01-06 22:26:47] Some components changed by more than 1% on the last iteration
Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Some components
changed by more than 1% on the last iteration
WARN [2025-01-06 22:26:47] Some components changed by more than 1% on the last iteration
Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Some components
changed by more than 1% on the last iteration
current.asrt <- as.asrtests(current.asr, wald.tab = NULL, test.summary = current.asrt$test.summary,
                            IClikelihood = "full", label = "Max model & Unbound components")
current.asrt <- rmboundary(current.asrt)</pre>
summary(current.asrt$asreml.obj)$varcomp
                                                  z.ratio bound %ch
                          component
                                      std.error
                     -2462.3785858 1.191435e+03 -2.066734
                                                              U 0.2
Rep
Rep:Row
                      5012.4021415 3.396848e+03 1.475604
                                                              U 0.1
Rep:Column
                       920.5936391 1.704008e+03 0.540252
                                                              U 1.1
                                                              P 0.1
units
                      5964.9099377 1.608792e+03 3.707695
Row:Column!R
                     46690.4620387 2.731906e+04 1.709080
                                                              P 0.0
                         0.8152180 9.988929e-02 8.161216
                                                              U 0.1
Row:Column!Row!cor
Row:Column!Column!cor
                         0.8857252 7.487875e-02 11.828793
                                                              U 0.0
print(current.asrt, which = "testsummary")
```

Sequence of model investigations for yield

ASReml Version 4.2 06/01/2025 22:26:46

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

```
terms DF denDF p AIC BIC action
Maximal model 26 6 NA 1646.129 1742.470 Starting model
Rep 1 NA NA 1646.129 1742.470 Boundary
Max model & Unbound components 26 7 NA 1647.200 1746.551 Starting model
```

```
print(current.asrt, which = "pseudoanova")
```

Pseudo-anova table for fixed terms

Wald tests for fixed effects. Response: yield

```
Df denDF F.inc Pr (Intercept) 1 1.7 153.400 0.0115 WithinColPairs 1 15.6 2.543 0.1308 Variety 24 76.1 10.110 0.0000
```

Now the Rep component estimate is negative.

The test.summary output has been extended, by supplying the previous test.summary to as.asrtests, to show that there is a new starting model. The pseudo-anova table shows that Varieties are highly significant (p < 0.001)

2. Perform a series of hypothesis tests to select a linear mixed model for the data

The hypothesis tests in this section are Wald tests for fixed terms, with denominator degrees of freedom calculated using the Kenward-Rogers adjustment (Kenward and Rogers (1997), and Restricted Maximum Likelihood Ratio Tests (REMLRT) for random terms.

Check the term for within Column pairs (a post hoc factor)

The information criteria based on the full likelihood (Verbyla, 2019) is also included in the test.summary stored in the asrtests object.

WARN [2025-01-06 22:26:48] Some components changed by more than 1% on the last iteration

Warning in asreml(fixed = yield \sim Variety, random = \sim Rep/(Row + Column) + : Some components changed by more than 1% on the last iteration

WARN [2025-01-06 22:26:49] Some components changed by more than 1% on the last iteration

Warning in asreml(fixed = yield \sim Variety, random = \sim Rep/(Row + Column) + : Some components changed by more than 1% on the last iteration

```
print(current.asrt)
```

Summary of the fitted variance parameters for yield

	component	std.error	z.ratio	${\tt bound}$	%ch
Rep	-2392.1616314	1.199592e+03	-1.9941460	U	0.4
Rep:Row	5033.2850607	3.408523e+03	1.4766764	U	0.2
Rep:Column	760.1498938	1.617038e+03	0.4700879	U	2.5
units	5929.0518909	1.609478e+03	3.6838361	P	0.0
Row:Column!R	45940.6913909	2.634982e+04	1.7434920	P	0.0
Row:Column!Row!cor	0.8101561	9.995026e-02	8.1055925	U	0.1
Row:Column!Column!cor	0.8846454	7.504265e-02	11.7885681	U	0.0

Pseudo-anova table for fixed terms

Wald tests for fixed effects. Response: yield

```
Df denDF F.inc Pr (Intercept) 1 1.7 159.00 0.0112 Variety 24 76.8 10.27 0.0000
```

Sequence of model investigations for yield

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

		ter	rms I	ΟF	denDF	р	AIC	BIC	action
1		Maximal mod	del 2	26	6.0	NA	1646.129	1742.470	Starting model
2		F	Rep	1	NA	NA	1646.129	1742.470	Boundary
3	Max model &	Unbound componer	nts 2	26	7.0	NA	1647.200	1746.551	Starting model
4		WithinColPai	irs	1	15.6	0.1308	1645.326	1741.666	Dropped

It is clear in the call to testranfix that the model is being changed by dropping the withinColPairs term, which could also be achieved using update.asreml. However, an asremlPlus model-changing function operates on an asrtests object, that includes an asreml object, and, except for changeTerms.asrtests, results in an asrtests object that may contain the changed model or the supplied model depending on the results of hypothesis tests or comparisons of information criteria. In addition, the result of the test or comparison will be added to a test.summary data.frame stored in the new asrtests object and, if the model was changed, the wald.tab in the new asrtests object will have been updated for the new model.

In this case, as can be seen from the summary of current.asrt after the call, the *p*-value for the withinColPairs was greater than 0.05 and so now the model stored in current.asrt does not include withinColPAirs. The wald.tab has been updated for the new model.

Test the nugget term

The nugget term represents non-spatial variance, such as random plot and measurement error. It is fitted using the asreml reserved word units.

```
current.asrt <- testranfix(current.asrt, "units", positive=TRUE, IClikelihood = "full")</pre>
```

WARN [2025-01-06 22:26:50] Some components changed by more than 1% on the last iteration

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + Rep:Row + Rep:Column, : Some components changed by more than 1% on the last iteration

WARN [2025-01-06 22:26:50] Some components changed by more than 1% on the last iteration

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + Rep:Row + Rep:Column, : Some components changed by more than 1% on the last iteration

Test Row autocorrelation

We begin testing the autocorrelation by dropping the Row autocorrelation. Because of messages about the instability of the fit, iterate.asrtests is used to execute extra iterations of the fitting process.

WARN [2025-01-06 22:26:53] Some components changed by more than 1% on the last iteration

Warning in asreml(fixed = yield ~ Variety, random = ~Rep/(Row + Column) + : Some components changed by more than 1% on the last iteration

WARN [2025-01-06 22:26:53] Some components changed by more than 1% on the last iteration

Warning in asreml(fixed = yield ~ Variety, random = ~Rep/(Row + Column) + : Some components changed by more than 1% on the last iteration

```
current.asrt <- iterate(current.asrt)</pre>
```

WARN [2025-01-06 22:26:54] Some components changed by more than 1% on the last iteration

Warning in asreml(fixed = yield ~ Variety, random = ~Rep/(Row + Column) + : Some components changed by more than 1% on the last iteration

WARN [2025-01-06 22:26:54] Some components changed by more than 1% on the last iteration

Warning in asreml(fixed = yield \sim Variety, random = \sim Rep/(Row + Column) + : Some components changed by more than 1% on the last iteration

Test Column autocorrelation (depends on whether Row autocorrelation retained)

The function getTestPvalue is used to get the p-value for the Row autocorrelation test. If it is significant then the Column autocorrelation is tested by dropping the Column autocorrelation, while retaining the Row autocorrelation. Otherwise the model with just Row autocorrelation, whose fit is returned via current.asrt after the test, is compared to one with no autocorrelation.

Warning in DFdiff(bound.h1, bound.h0, DF = DF, bound.exclusions = bound.exclusions): There were a total The following bound terms occur in only one of the models compared and so were discounted:

Row:Column!Row!cor

Output the results

```
print(current.asrt)
```

Summary of the fitted variance parameters for yield

```
component
                                      std.error
                                                   z.ratio bound %ch
                      -2392.1616314 1.199592e+03 -1.9941460
                                                               U 0.4
Rep
Rep:Row
                      5033.2850607 3.408523e+03 1.4766764
                                                               U 0.2
Rep:Column
                       760.1498938 1.617038e+03 0.4700879
                                                               U 2.5
                      5929.0518909 1.609478e+03 3.6838361
units
                                                               P 0.0
Row:Column!R
                     45940.6913909 2.634982e+04 1.7434920
                                                               P 0.0
Row:Column!Row!cor
                         0.8101561 9.995026e-02 8.1055925
                                                               U 0.1
Row:Column!Column!cor
                         0.8846454 7.504265e-02 11.7885681
                                                               U 0.0
```

Pseudo-anova table for fixed terms

Wald tests for fixed effects.
Response: yield

Df denDF F.inc Pr
(Intercept) 1 1.7 159.00 0.0112
Variety 24 76.8 10.27 0.0000

Sequence of model investigations for yield

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

```
terms DF denDF
                                                                BIC
                                                       AIC
                                                                            action
                                               р
                   Maximal model 26
1
                                      6.0
                                              NA 1646.129 1742.470 Starting model
2
                                       NA
                                              NA 1646.129 1742.470
                                                                          Boundary
                             Rep 1
3 Max model & Unbound components 26
                                      7.0
                                              NA 1647.200 1746.551 Starting model
                  WithinColPairs 1 15.6 0.1308 1645.326 1741.666
                                                                           Dropped
5
                                       NA 0.0006 1645.326 1741.666
                           units 1
                                                                          Retained
             Row autocorrelation 1
6
                                       NA 0.0000 1645.326 1741.666
                                                                         Unswapped
             Col autocorrelation 2
                                       NA 0.0000 1645.326 1741.666
                                                                         Unswapped
printFormulae(current.asrt$asreml.obj)
#### Formulae from asreml object
fixed: yield ~ Variety
random: ~ Rep/(Row + Column) + units
residual: ~ ar1(Row):ar1(Column)
print(R2adj(current.asrt$asreml.obj, include.which.random = ~ .))
ASReml Version 4.2 06/01/2025 22:26:56
          LogLik
                        Sigma2
                                   \mathsf{DF}
                                          wall
1
       -694.6149
                      45873.83
                                  125
                                        22:26:56
2
       -694.6149
                      45868.25
                                  125
                                        22:26:56
[1] 44.62078
attr(,"fixed")
<environment: 0x0000017d7f6b6ac8>
attr(,"random")
```

The test.summary shows is that the model with Row and without Column autocorrelation failed to converge. The asreml.obj in current.asrt contains the model selected by the selection process, which has been printed using printFormulae.asrtests. It is clear that no changes were made to the variance terms. The adjusted R^2 value shows that the fixed and random terms in the fitted model account for 45% of the total variation in the yield.

3. Diagnostic checking using residual plots and variofaces

Get current fitted asreml object and update to include standardized residuals

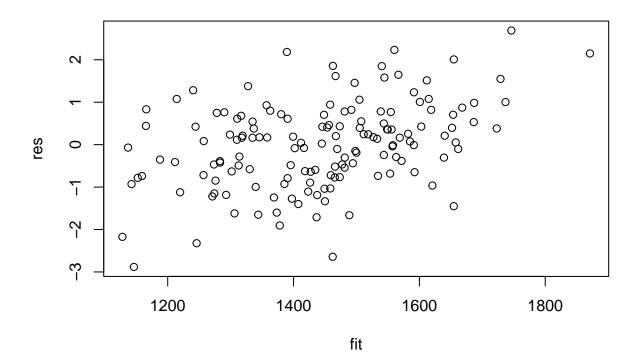
```
current.asr <- current.asrt$asreml.obj</pre>
current.asr <- update(current.asr, aom=TRUE)</pre>
ASReml Version 4.2 06/01/2025 22:26:57
                                     DF
          LogLik
                         Sigma2
                                             wall
1
       -694.6149
                       45873.83
                                    125
                                           22:26:57
 2
       -694.6149
                       45868.25
                                    125
                                           22:26:57
 3
       -694.6149
                                    125
                       45854.78
                                           22:26:57
```

```
Wheat.dat$res <- residuals(current.asr, type = "stdCond")
Wheat.dat$fit <- fitted(current.asr)</pre>
```

Do diagnostic checking

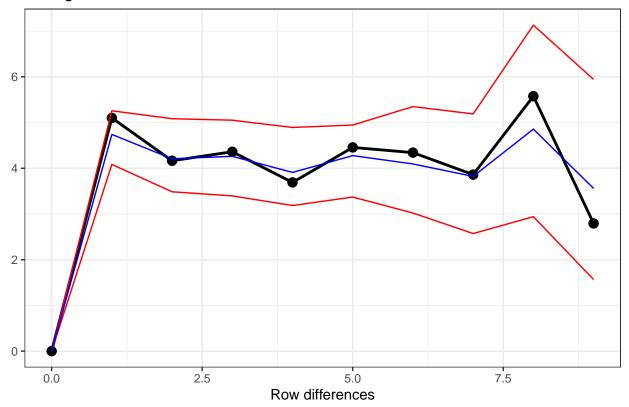
Do residuals-versus-fitted values plot

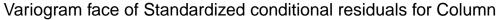
```
with(Wheat.dat, plot(fit, res))
```

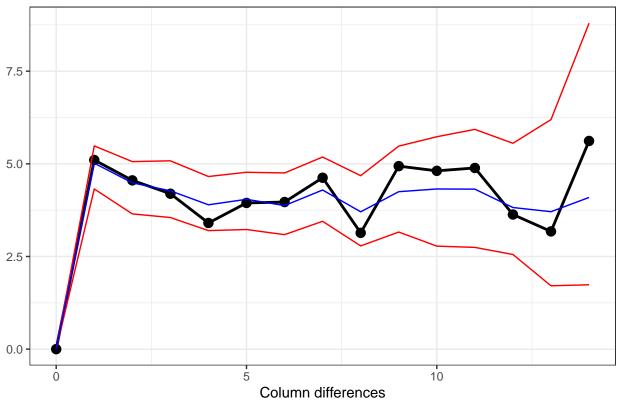


Plot variofaces

Variogram face of Standardized conditional residuals for Row



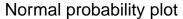


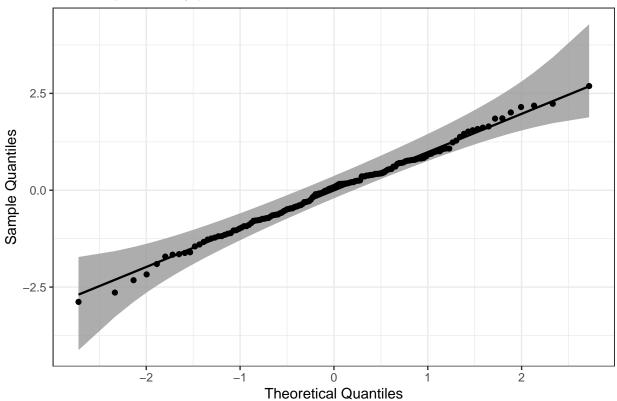


The variofaces are the lag 1 plots of the sample semivariogram with simulated confidence envelopes (Stefanova et al., 2009).

Plot normal quantile plot

The plot is obtained using the ggplot function with extensions available from the qqplotr package (Almeida, A., Loy, A. and Hofmann, H., 2023).





4. Prediction production and presentation

Get Variety predictions and all pairwise prediction differences and p-values

Predictions for yield from Variety

Notes:

- The predictions are obtained by averaging across the hypertable calculated from model terms constructed solely from factors in the averaging and classify sets.
- Use 'average' to move ignored factors into the averaging set.
- The ignored set: Rep,Row,Column,units

- Variety is included in this prediction
- (Intercept) is included in this prediction
- units is ignored in this prediction

12

13 14

15 16

17

18

19

20

21

22

23

	Variety	predicted value	gtandard	arror	upper.halfLeastSignificant.limit
1	10	1168.989).4776	1228.315
2	1	1242.750		0.8112	1302.076
3	9	1257.137		9.9716	1316.463
4	16	1285.719		9.9407	1345.045
5	14			9.9234	1352.853
6	23	1313.653	120	.2937	1372.979
7	11	1322.159	120	.1971	1381.485
8	7	1374.446	120	.2415	1433.773
9	3	1394.070	120	.4040	1453.396
10	4	1410.980	120	.1063	1470.306
11	12	1444.557	120	.6042	1503.883
12	8	1453.396	120	.5948	1512.722
13	15	1458.383	120	.4354	1517.709
14	5	1473.782	120	.4462	1533.108
15	17	1487.828	120	.2904	1547.154
16	6	1498.294	120	.1196	1557.621
17	21	1517.121	120	.2270	1576.447
18	2	1520.466		6330	1579.792
19	24	1533.768		.3003	1593.095
20	18	1541.148		.3671	1600.474
21	25	1575.795		.5149	1635.121
22				.3289	1669.808
23	13	1610.761		.4582	1670.088
24	20	1627.971		.2336	1687.297
25	19	1652.992		.3443	1712.318
	lower.ha	${ t alfLeastSignific}$			
1			1109.662	Estin	
2			1183.424	Estin	
3			1197.811	Estin	
4			1226.392	Estin	
5			1234.200	Estin	
6			1254.327	Estin	
7			1262.832	Estin	
8			1315.120	Estin	
9			1334.743	Estin	
10			1351.654	Estin	
11			1385.231	Estin	парте

1394.070 Estimable 1399.057 Estimable

1414.456 Estimable 1428.501 Estimable

1438.968 Estimable

1457.795 Estimable

1461.140 Estimable

1474.442 Estimable

1481.822 Estimable

1516.468 Estimable

1551.156 Estimable

1551.435 Estimable

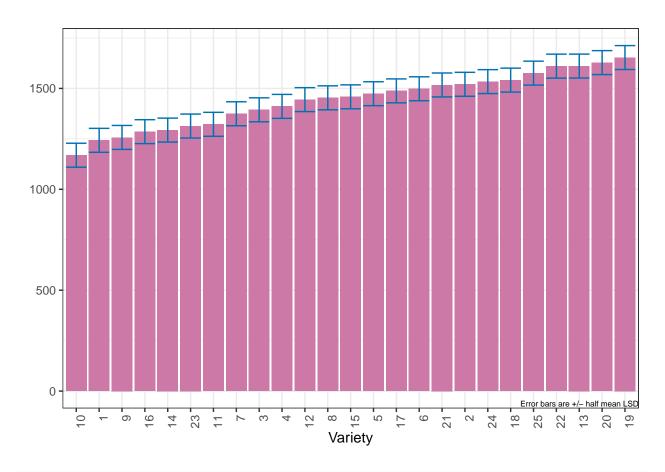
We have set error.intervals to halfLeast so that the limits for so that the limits for each prediction \pm (0.5 LSD) are calculated. When these are plotted overlapping error bars indicate predictions that are not significant, while those that do not overlap are significantly different (Snee, 1981).

Also set was sortFactor, so that the results would be ordered for the values of the predictions for Variety.

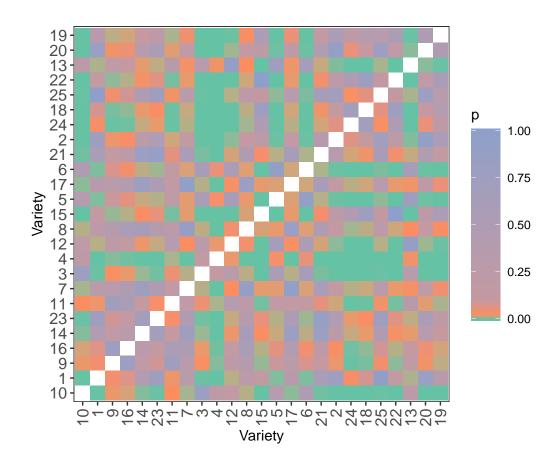
The function predictPlus returns an alldiffs object, a list consisting of the following components:

- predictions: the predictions, their standard errors and error intervals;
- vcov: the variance matrix of the predictions;
- differences: all pairwise differences between the predictions,
- p.differences: p-values for all pairwise differences between the predictions;
- sed: the standard errors of all pairwise differences between the predictions;
- LSD: the mean, minimum and maximum LSDs.

Plot the Variety predictions, with halfLSD intervals, and the p-values



plotPvalues(Var.diffs)



References

Almeida, A., Loy, A. and Hofmann, H. (2023) qqplotr: Quantile-Quantile plot extensions for 'ggplot2', Version 0.0.6. https://cran.r-project.org/package=qqplotr/ or https://github.com/aloy/qqplotr/.

Brien, C. J. (2025) asremlPlus: Augments ASReml-R in fitting mixed models and packages generally in exploring prediction differences. Version 4.4.44. https://cran.r-project.org/package=asremlPlus/ or http://chris.brien.name/rpackages/.

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J. and Thompson, R. (2023). ASReml-R Reference Manual Version 4.2. VSN International Ltd, https://asreml.kb.vsni.co.uk/.

Gilmour, A. R., Thompson, R., & Cullis, B. R. (1995). Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics*, **51**, 1440–1450.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983-997.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). SAS for Mixed Models (2nd ed.). Cary, N.C.: SAS Press.

R Core Team (2025) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Snee, R. D. (1981). Graphical Display and Assessment of Means. *Biometrics*, 37, 835–836.

Stefanova, K. T., Smith, A. B. & Cullis, B. R. (2009) Enhanced diagnostics for the spatial analysis of field trials. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 392–410.

Verbyla, A. P. (2019). A note on model selection using information criteria for general linear models estimated using REML. Australian & New Zealand Journal of Statistics, $\bf 61$, $\bf 39\text{-}50.https://doi.org/10.1111/anzs.12254/.$