Using lm with asremlPlus for the Ladybird example from Welham et al. (2014)

Chris Brien

18 October, 2023

# Introduction

This vignette shows how to use asremlPlus (Brien, 2023a), and dae (Brien, 2023b), for exploring and presenting predictions from a linear mixed model analysis, the predictions having been produced using lmerTest (Kuznetsova et al., 2017), lm and emmeans (Lenth, 2023). Here, asremlPlus, dae, lmerTest and emmeans are packages for the R Statistical Computing environment (R Core Team, 2023) and lm is available from stats and is included in R.

The context is a three-factor factorial experiment on ladybirds (Welham et al., 2014, Example 8.2) that aims to answer the question "Will ladybirds transfer fungus to aphids on plants?" The experiment consists of 2 runs of 36 containers, each with a plant and aphids. There are three factors that results in 12 treatments: Host plant (beans, trefoil), infected Cadavers (5, 10, 20), Ladybird (-, +). Ther are randomized to the containers within a run so that each is replicated 3 times within a run. The respose to be analysed is the logit of the proportion of live aphids that were infected.

### Initialize

```
library(knitr)
opts_chunk$set("tidy" = FALSE, comment = NA)
suppressMessages(library(lmerTest))
packageVersion("lmerTest")

## [1] '3.1.3'
suppressMessages(library(emmeans))
packageVersion("emmeans")

## [1] '1.8.8'
suppressMessages(library(asremlPlus))
packageVersion("asremlPlus")

## [1] '4.4.16'
suppressMessages(library(dae))
packageVersion("dae")

## [1] '3.2.19'
options(width = 95, show.signif.stars = FALSE)
```

#### Get data available in asremlPlus

```
data("Ladybird.dat")
```

# Do an ANOVA of logits

```
Error: Run

Df Sum Sq Mean Sq F value Pr(>F)
Residuals 1 0.06766 0.06766
```

Error: Run:Plant

	Df	Sum Sq	Mean Sq	${\tt F} \ {\tt value}$	Pr(>F)
Host	1	13.599	13.599	59.172	1.82e-10
Cadavers	2	17.027	8.514	37.044	3.78e-11
Ladybird	1	11.091	11.091	48.257	3.33e-09
Host:Cadavers	2	0.308	0.154	0.670	0.5158
Host:Ladybird	1	0.228	0.228	0.992	0.3234
Cadavers:Ladybird	2	1.735	0.867	3.774	0.0287
Host:Cadavers:Ladybird	2	0.200	0.100	0.435	0.6493
Residuals	59	13.560	0.230		

The anova table gives the F-tests for the three-factor effects and interactions. Note the Residuals Mean Sq value for Run:Plant of 0.230. Also, it is clear that the Run component is negative, given that the Residuals Mean Sq value for Run is less than that for Run:Plant; it is (0.06766-0.230)/36). From the table it is seen that the only significant interaction is Cadavers:Ladybird and that the Host main effect is significant.

# Use lmerTest and lm to analyse the logits

# Mixed model analysis of logits

```
m1.lmer <- lmerTest::lmer(logitP ~ Host*Cadavers*Ladybird + (1|Run),
                          data=Ladybird.dat)
boundary (singular) fit: see help('isSingular')
summary(m1.lmer)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: logitP ~ Host * Cadavers * Ladybird + (1 | Run)
   Data: Ladybird.dat
REML criterion at convergence: 102.8
Scaled residuals:
             10 Median
                             3Q
   Min
                                    Max
-1.9633 -0.5217 0.1360 0.5789 2.1896
Random effects:
Groups Name
                     Variance Std.Dev.
```

```
Run
          (Intercept) 0.0000
                              0.0000
                     0.2271
                              0.4766
Residual
Number of obs: 72, groups: Run, 2
Fixed effects:
                                                            df t value Pr(>|t|)
                                 Estimate Std. Error
(Intercept)
                                            0.194560 60.000000 -8.240 1.91e-11
                                -1.603097
Hosttrefoil
                                            0.275149 60.000000 -3.164 0.00244
                                -0.870675
Cadavers10
                                 0.564771
                                            0.275149 60.000000
                                                                 2.053
                                                                        0.04448
Cadavers20
                                 0.919229
                                            0.275149 60.000000
                                                                 3.341 0.00144
Ladybird+
                                 0.547710
                                            0.275149 60.000000
                                                                1.991 0.05109
                                            0.389120 60.000000 -0.547
Hosttrefoil:Cadavers10
                                -0.212735
                                                                        0.58661
Hosttrefoil:Cadavers20
                                -0.120410
                                            0.389120 60.000000 -0.309 0.75806
Hosttrefoil:Ladybird+
                                 0.073153
                                            0.389120 60.000000
                                                                 0.188 0.85151
Cadavers10:Ladybird+
                                -0.040048
                                            0.389120 60.000000 -0.103 0.91837
Cadavers20:Ladybird+
                                 0.414204
                                            0.389120 60.000000
                                                                 1.064
                                                                        0.29138
                                                                 0.010 0.99177
Hosttrefoil:Cadavers10:Ladybird+
                                 0.005698
                                            0.550299 60.000000
Hosttrefoil:Cadavers20:Ladybird+ 0.449979
                                            0.550299 60.000000
                                                                 0.818 0.41676
Correlation of Fixed Effects:
            (Intr) Hsttrf Cdvr10 Cdvr20 Ldybr+ Hs:C10 Hs:C20 Hst:L+ C10:L+ C20:L+ H:C10:
Hosttrefoil -0.707
Cadavers10 -0.707 0.500
Cadavers20 -0.707 0.500 0.500
Ladybird+
           -0.707 0.500 0.500 0.500
Hsttrfl:C10 0.500 -0.707 -0.707 -0.354 -0.354
Hsttrfl:C20 0.500 -0.707 -0.354 -0.707 -0.354
                                               0.500
Hsttrfl:Ld+ 0.500 -0.707 -0.354 -0.354 -0.707
                                               0.500
                                                      0.500
Cdvrs10:Ld+ 0.500 -0.354 -0.707 -0.354 -0.707
                                              0.500 0.250
                                                            0.500
Cdvrs20:Ld+ 0.500 -0.354 -0.354 -0.707 -0.707 0.250 0.500 0.500 0.500
Hstt:C10:L+ -0.354  0.500  0.500  0.250  0.500 -0.707 -0.354 -0.707 -0.707 -0.354
```

As expected the Run component is bound at zero, leading to a singular model. This results in a change in the estimate of the residual variance to 0.227. To allow for a negative estimate we will redo the analysis with Run fixed, because with lme4 (lmerTest) one cannot unconstrain the Run component to allow it to be negative. As Littell et al. (2006, p.150) say

Hstt:C20:L+ -0.354 0.500 0.250 0.500 0.500 -0.354 -0.707 -0.707 -0.354 -0.707 0.500

if you do not set the negative variance component estimate to zero, but allow it to remain negative, you get better control over Type I error and, for cases of negative wholeplot error variance estimates, greater power. Therefore, this is the recommended procedure.

#### Analyse with Reps fixed using 1m to make the analysis equivalent to ANOVA

The function 1m has to be used because there are no random terms; 1me4 cannot be used because it requires at least one random term.

Analysis of Variance Table

optimizer (nloptwrap) convergence code: 0 (OK) boundary (singular) fit: see help('isSingular')

Response: logitP

```
Df Sum Sq Mean Sq F value
                                                   Pr(>F)
Run
                       1 0.0677 0.0677 0.2944
                                                  0.58946
Host
                       1 13.5992 13.5992 59.1720 1.815e-10
Cadavers
                       2 17.0274 8.5137 37.0444 3.784e-11
Ladybird
                       1 11.0907 11.0907 48.2571 3.329e-09
Host:Cadavers
                       2 0.3078 0.1539 0.6695
                                                  0.51579
Host:Ladybird
                       1 0.2279 0.2279 0.9916
                                                  0.32341
Cadavers:Ladybird
                       2 1.7349 0.8675 3.7744
                                                  0.02867
Host:Cadavers:Ladybird 2 0.1999 0.1000 0.4350
                                                  0.64932
Residuals
                      59 13.5596 0.2298
```

Now the Run: Plant variance estimate is equal to that for the Residuals Mean Sq for Run: Plant from the anova table.

# Obtain the marginality matrix for the fixed terms

The pstructure function from the dae package (Brien, 2023b) produce the marginality matrix for a formula as a side effect and we take advantage of that to obtain the matrix required here.

	Host	Cadavers	Host:Cadavers	Ladybird	${\tt Host:Ladybird}$	Cadavers:Ladybird
Host	1	0	1	0	1	0
Cadavers	0	1	1	0	0	1
Host:Cadavers	0	0	1	0	0	0
Ladybird	0	0	0	1	1	1
Host:Ladybird	0	0	0	0	1	0
Cadavers:Ladybird	0	0	0	0	0	1
Host:Cadavers:Ladybird	0	0	0	0	0	0

Host:Cadavers:Ladybird
Host
Cadavers
1
Cadavers
1
Host:Cadavers
1
Ladybird
1
Host:Ladybird
1
Cadavers:Ladybird
1
Host:Cadavers:Ladybird
1
Host:Cadavers:Ladybird
1

This marginality matrix is interpreted by taking a row term and noting that it is marginal to any column term with a one in this row.

#### Choose marginality-compliant model

#### Sequence of model investigations

```
terms DF denDF p action
1 Host:Cadavers:Ladybird 2 59 0.6493 Nonsignificant
2 Cadavers:Ladybird 2 59 0.0287 Significant
```

The chooseModel function produces a list with components sig.terms, a list with the terms in the marginality-compliant model, and choose.summary, a data frame that details the tests performed in choosing the model. Note that chooseModel does not test the main effects for Cadavers or Ladybird, because these are marginal to the significant two-factor interaction Cadavers:Ladybird.

#### Form formula for selected model

```
chosen.mod <- paste(unlist(chosen$sig.terms), collapse = " + ")
(chosen.mod <- as.formula(paste("~", chosen.mod)))

~Cadavers:Ladybird + Host</pre>
```

### Form predictions that conform to the chosen model

Use emmeans to get the predictions and associated statistics for the full model.

```
HCL.emm <- emmeans::emmeans(m1.lmer, specs = ~ Host:Cadavers:Ladybird)
HCL.preds <- summary(HCL.emm)
den.df <- min(HCL.preds$df)
HCL.vcov <- vcov(HCL.emm)</pre>
```

Setting the specs argument to Host:Ladybird:Cadavers requests predictions for all combinations of the three factors.

#### Modify HCL.preds to be compatible with a predictions.frame

Basically, this is an exercise in renaming the columns in the data.frame containing the predictions.

#### Form an alldiffs object with predictions obtained with emmeans

[7] "lower.Confidence.limit" "upper.Confidence.limit" "est.status"

The functions allDifferences is used to form the alldiffs.obj that contains a predictions component, along with components related to pairwise comparisons. The predictions component contains upper and lower confidence limits produced by emmeans. The tdf is supplied so that it can be used to get the degrees of freedom for the t-value to be used in calculating the error intervals.

#### Transform the prediction to conform to chosen model

The linTransform function is used to obtain estimated marginal means (emm) that conform to the chosen model. Because we would prefer error intervals based on  $\pm 0.5 LSD$ , the error intervals argument has been set to "halfLeast", the LSDtype argument to "factor.combination" and the LSDby argument to "Host" so that the average LSD will be calculated for each Host. This necessary because, under the chosen model, the LSDs differ between Hosts. It results in lower.halfLeastSignificant.limit and upper.halfLeastSignificant.limit replacing the limits based on the confidence intervals in the predictions component of the resulting alldiffs object.

```
Joining with `by = join_by(fac.comb)`
Joining with `by = join_by(Host)`
```

#### Predictions for transform(s) from Host:Ladybird:Cadavers

The original predictions, obtained as described below, have been linearly transformed to form estimated marginal means.

```
Host Ladybird Cadavers predicted.value standard.error
                                                                 df
1
      bean
                            5
                                   -1.6038338
                                                    0.1485977 47.2
2
                                   -1.1454308
                                                    0.1485977 47.2
      bean
                           10
3
      bean
                           20
                                   -0.7448097
                                                    0.1485977 47.2
4
      bean
                            5
                                   -1.0195475
                                                    0.1485977 47.2
5
                           10
                                   -0.5983440
                                                    0.1485977 47.2
      bean
6
      bean
                           20
                                    0.4786704
                                                    0.1485977 47.2
7
   trefoil
                            5
                                   -2.4730339
                                                    0.1485977 47.2
                                                    0.1485977 47.2
8
  trefoil
                           10
                                   -2.0146309
                           20
9 trefoil
                                   -1.6140098
                                                    0.1485977 47.2
10 trefoil
                            5
                                   -1.8887476
                                                    0.1485977 47.2
11 trefoil
                           10
                                   -1.4675441
                                                    0.1485977 47.2
12 trefoil
                           20
                                   -0.3905297
                                                    0.1485977 47.2
```

upper.halfLeastSignificant.limit lower.halfLeastSignificant.limit est.status 1 -1.4081535-1.7995140 Estimable 2 -0.9497506 -1.3411111 Estimable 3 -0.5491295 -0.9404900 Estimable 4 -0.8238673 -1.2152278 Estimable 5 -0.4026637 -0.7940242 Estimable 0.2829901 Estimable 6 0.6743507 7 -2.2773537 -2.6687142 Estimable 8 -1.8189507 -2.2103112 Estimable 9 -1.4183296 -1.8096901 Estimable 10 Estimable -1.6930674 -2.0844279 11 -1.2718638 -1.6632243 Estimable 12 -0.1948495 -0.5862100 Estimable

```
LSD values
minimum LSD = 0.3913605 0.3913605
```

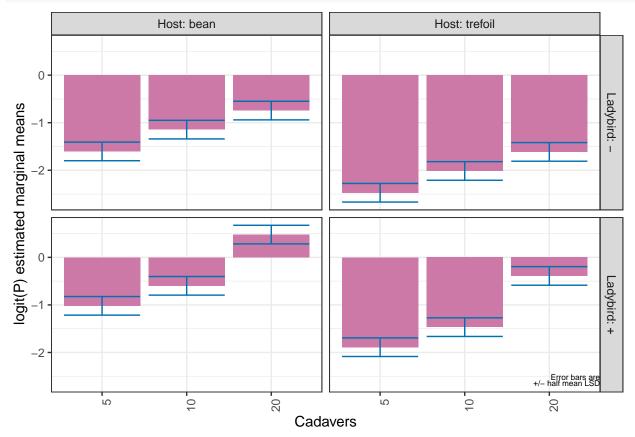
maximum LSD = 0.3913605 0.3913605

mean LSD = 0.3913605 0.3913605

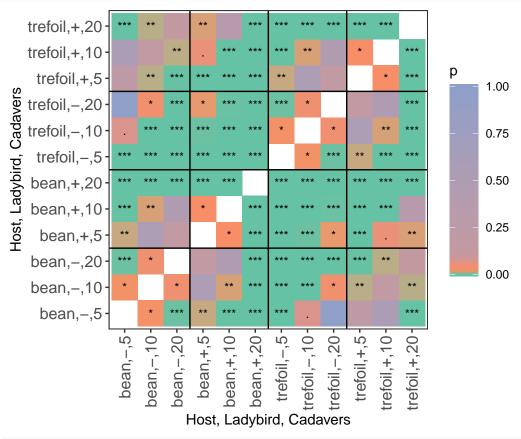
(sed range / mean sed = 3.49e-14 3.45e-14)

# Plot the predictions

The function plotPredictions uses ggplot to produce the plot and the ggplotFuncs argument allows the addition of ggplot functions to modify the plot. In this case, the facet.grid function is respecified to include prepender functions that modify the labels of the facets to include the factor names. Note the the error bars in the plots are of  $\pm 0.5 LSD$  so that pairs of prediction with nonoverlapping bars are significantly different (Snee, 1981).



#### plotPvalues(diffs, factors.per.grid = 1, show.sig = TRUE)



# options(width = 90) print(diffs\$sed)

```
bean,-,5 bean,-,10 bean,-,20 bean,+,5 bean,+,10 bean,+,20 trefoil,-,5
bean, -,5
                    NA 0.1945600 0.1945600 0.1945600 0.1945600 0.1945600
                                                                            0.1123293
bean, -, 10
                              NA 0.1945600 0.1945600 0.1945600 0.1945600
             0.1945600
                                                                            0.2246586
             0.1945600 0.1945600
                                         NA 0.1945600 0.1945600 0.1945600
bean, -, 20
                                                                            0.2246586
bean,+,5
             0.1945600 0.1945600 0.1945600
                                                   NA 0.1945600 0.1945600
                                                                            0.2246586
bean,+,10
             0.1945600 0.1945600 0.1945600 0.1945600
                                                             NA 0.1945600
                                                                            0.2246586
             0.1945600 0.1945600 0.1945600 0.1945600 0.1945600
                                                                            0.2246586
bean,+,20
                                                                       NA
trefoil,-,5 0.1123293 0.2246586 0.2246586 0.2246586 0.2246586 0.2246586
                                                                                    NA
trefoil,-,10 0.2246586 0.1123293 0.2246586 0.2246586 0.2246586 0.2246586
                                                                            0.1945600
trefoil,-,20 0.2246586 0.2246586 0.1123293 0.2246586 0.2246586 0.2246586
                                                                            0.1945600
trefoil,+,5 0.2246586 0.2246586 0.1123293 0.2246586 0.2246586
                                                                             0.1945600
trefoil,+,10 0.2246586 0.2246586 0.2246586 0.1123293 0.2246586
                                                                            0.1945600
trefoil,+,20 0.2246586 0.2246586 0.2246586 0.2246586 0.2246586 0.1123293
                                                                             0.1945600
             trefoil,-,10 trefoil,-,20 trefoil,+,5 trefoil,+,10 trefoil,+,20
bean, -,5
                0.2246586
                             0.2246586
                                          0.2246586
                                                       0.2246586
                                                                    0.2246586
bean, -, 10
                0.1123293
                             0.2246586
                                          0.2246586
                                                       0.2246586
                                                                    0.2246586
bean, -, 20
                0.2246586
                             0.1123293
                                          0.2246586
                                                       0.2246586
                                                                    0.2246586
bean,+,5
                0.2246586
                             0.2246586
                                          0.1123293
                                                       0.2246586
                                                                    0.2246586
                0.2246586
                             0.2246586
                                                       0.1123293
bean,+,10
                                          0.2246586
                                                                    0.2246586
bean, +, 20
                0.2246586
                             0.2246586
                                          0.2246586
                                                       0.2246586
                                                                    0.1123293
                             0.1945600
trefoil,-,5
                0.1945600
                                          0.1945600
                                                       0.1945600
                                                                    0.1945600
trefoil,-,10
                             0.1945600
                                         0.1945600
                                                       0.1945600
                                                                    0.1945600
                       NA
```

```
0.1945600
trefoil,-,20
                0.1945600
                                    NA
                                                      0.1945600
                                                                    0.1945600
                             0.1945600
                                                      0.1945600
                                                                    0.1945600
trefoil,+,5
                0.1945600
                                                NA
                                         0.1945600
trefoil,+,10
                0.1945600
                             0.1945600
                                                                    0.1945600
trefoil,+,20
                             0.1945600
                                         0.1945600
                0.1945600
                                                      0.1945600
                                                                           NA
```

# Perform the analysis with just the selected model fitted

The model with nonsignificant fixed terms dropped is obtained in order to compare it with the fit when they are retained and the etimated marginal means for the chosen model are obtained.

```
m.sig.lm <- lm(logitP ~ Run + Cadavers*Ladybird + Host,
                         data=Ladybird.dat)
(aov.tab <- anova(m.sig.lm))</pre>
Analysis of Variance Table
Response: logitP
                  Df Sum Sq Mean Sq F value
                   1 0.0677 0.0677 0.3029
Run
                                                0.58398
Cadavers
                   2 17.0274 8.5137 38.1160 1.255e-11
                   1 11.0907 11.0907 49.6531 1.542e-09
Ladybird
                   1 13.5992 13.5992 60.8836 7.179e-11
Cadavers:Ladybird 2 1.7349 0.8675 3.8836
                                                0.02559
Residuals
                  64 14.2952 0.2234
HCL.emm <- emmeans::emmeans(m.sig.lm, specs = ~ Host:Cadavers:Ladybird)</pre>
HCL.preds <- summary(HCL.emm)</pre>
den.df <- min(HCL.preds$df)</pre>
HCL.vcov <- vcov(HCL.emm)</pre>
HCL.preds <- as.predictions.frame(HCL.preds, predictions = "emmean",</pre>
                                   se = "SE", interval.type = "CI",
                                   interval.names = c("lower.CL", "upper.CL"))
diffs.red <- allDifferences(predictions = HCL.preds, classify = "Host:Ladybird:Cadavers",</pre>
                            vcov = HCL.vcov, tdf = den.df)
diffs.red <- redoErrorIntervals(diffs, error.intervals = "halfLeast",</pre>
                                 LSDtype = "factor.combination", LSDby = "Host")
Joining with `by = join_by(fac.comb)`
Joining with `by = join_by(Host)`
options(width = 90)
print(diffs.red$sed)
              bean,-,5 bean,-,10 bean,-,20 bean,+,5 bean,+,10 bean,+,20 trefoil,-,5
bean, -, 5
                    NA 0.1945600 0.1945600 0.1945600 0.1945600 0.1945600
                                                                             0.1123293
                              NA 0.1945600 0.1945600 0.1945600 0.1945600
bean,-,10
             0.1945600
                                                                             0.2246586
bean,-,20
                                         NA 0.1945600 0.1945600 0.1945600
             0.1945600 0.1945600
                                                                             0.2246586
             0.1945600 0.1945600 0.1945600
                                                   NA 0.1945600 0.1945600
bean,+,5
                                                                             0.2246586
                                                             NA 0.1945600
             0.1945600 0.1945600 0.1945600 0.1945600
bean,+,10
                                                                             0.2246586
bean,+,20
             0.1945600 0.1945600 0.1945600 0.1945600 0.1945600
                                                                             0.2246586
trefoil,-,5 0.1123293 0.2246586 0.2246586 0.2246586 0.2246586
```

0.1945600

0.1945600

0.1945600

0.1945600

0.1945600

trefoil,-,10 0.2246586 0.1123293 0.2246586 0.2246586 0.2246586 0.2246586

trefoil,-,20 0.2246586 0.2246586 0.1123293 0.2246586 0.2246586 0.2246586

trefoil,+,5 0.2246586 0.2246586 0.1123293 0.2246586 0.2246586

trefoil,+,10 0.2246586 0.2246586 0.2246586 0.1123293 0.2246586

trefoil,+,20 0.2246586 0.2246586 0.2246586 0.2246586 0.1123293

	trefoil,-,10	trefoil,-,20	trefoil,+,5	trefoil,+,10	trefoil,+,20
bean,-,5	0.2246586	0.2246586	0.2246586	0.2246586	0.2246586
bean,-,10	0.1123293	0.2246586	0.2246586	0.2246586	0.2246586
bean,-,20	0.2246586	0.1123293	0.2246586	0.2246586	0.2246586
bean,+,5	0.2246586	0.2246586	0.1123293	0.2246586	0.2246586
bean,+,10	0.2246586	0.2246586	0.2246586	0.1123293	0.2246586
bean,+,20	0.2246586	0.2246586	0.2246586	0.2246586	0.1123293
trefoil,-,5	0.1945600	0.1945600	0.1945600	0.1945600	0.1945600
trefoil,-,10	NA	0.1945600	0.1945600	0.1945600	0.1945600
trefoil,-,20	0.1945600	NA	0.1945600	0.1945600	0.1945600
trefoil,+,5	0.1945600	0.1945600	NA	0.1945600	0.1945600
trefoil,+,10	0.1945600	0.1945600	0.1945600	NA	0.1945600
trefoil,+,20	0.1945600	0.1945600	0.1945600	0.1945600	NA

# References

Brien, C. J. (2023a) asremlPlus: Augments ASReml-R in fitting mixed models and packages generally in exploring prediction differences. Version 4.4.16. https://cran.r-project.org/package=asremlPlus/ or http://chris.brien.name/rpackages/.

Brien, C. J. (2023b) dae: Functions useful in the design and ANOVA of experiments. Version 3.2-19. https://cran.r-project.org/package=dae/ or http://chris.brien.name/rpackages/.

Kuznetsova, A., Brockhoff, P. B. and Christensen, R. H. B. (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82, 1–26.

Lenth, R. V., Buerkner, P., Herve, M. J., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023) emmeans: *Estimated Marginal Means*, aka Least-Squares Means. Version 1.8.8. https://cran.r-project.org/package=emmeans/.

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

R Core Team (2023) 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.

Welham, S. J., Gezan, S. A., Clark, S. J., & Mead, A. (2014). Statistical Methods in Biology: Design and Analysis of Experiments and Regression. Boca Raton: Chapman and Hall/CRC.