Statistical modeling for phenotypic traits

ABER_StatisticalAnalysis

Elise

2024-06-09

Statistical modeling for phenotypic traits

The explanatory variable (X) refers to the genotype, which is a categorical variable. The response variables (Y) are the phenotypic data of the variables dataframe.

In this case, the variables are:

And the genotypes are:

```
print(variables)
```

```
## [1] "DW_shoot_g"  "FW_shoot_g"  "DW_root_g"  "FW_root_g"  "Leaf_number"
```

```
unique(endpoint$Genotype)
```

```
## [1] EPPN1_L EPPN2_L EPPN3_L EPPN4_L EPPN1_H EPPN2_H EPPN3_H EPPN4_H EPPN_T
## 9 Levels: EPPN_T EPPN1_H EPPN1_L EPPN2_H EPPN2_L EPPN3_H EPPN3_L ... EPPN4_L
```

1. First linear models

Firstly, we model the Y = X + r + c + e Where - Y is the phenotypic trait; - X the genotype (fixed effect); - r the row effect (fixed or random); - c the column effect (fixed or random);

```r
fit_models_fixed <- function(data, trait_name) {
  fixed_formula <- as.formula(paste(trait_name, "~ Genotype + Row + Column"))
  fixed_model <- lm(fixed_formula, data)
  print(paste("Summary for fixed effects model of", trait_name))
  print(summary(fixed_model))
  print(anova(fixed_model))
}

fit_models_random <- function(data, trait_name)  {
  random_formula <- as.formula(paste(trait_name, "~ Genotype + (1|Row) + (1|Column)"))
  random_model <- lmer(random_formula, data)
  print(paste("Summary for random effects model of", trait_name))
  print(summary(random_model))
  print(anova(random_model))
  print(ranova(random_model))
}

for (trait in variables) {
  fit_models_fixed(endpoint_clean, trait)
}
```

```
## [1] "Summary for fixed effects model of DW_shoot_g"
##
## Call:
## lm(formula = fixed_formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4190 -1.3959  0.1507  1.2243  6.7754
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      26.3453     2.6897   9.795 1.13e-09 ***
## GenotypeEPPN1_H  -1.4729     2.3686  -0.622 0.540164
## GenotypeEPPN1_L  -8.5607     2.2162  -3.863 0.000791 ***
## GenotypeEPPN2_H  -0.5504     2.5032  -0.220 0.827923
## GenotypeEPPN2_L  -6.1207     2.0123  -3.042 0.005794 **
## GenotypeEPPN3_H  -4.1444     2.0181  -2.054 0.051553 .
## GenotypeEPPN3_L  -5.9184     2.5771  -2.296 0.031091 *
## GenotypeEPPN4_H  -4.9984     2.1397  -2.336 0.028572 *
## GenotypeEPPN4_L -14.1098     1.9690  -7.166 2.69e-07 ***
## Row2             -6.8915     2.0305  -3.394 0.002494 **
## Row3             -6.6396     2.0230  -3.282 0.003269 **
## Row4             -6.6652     1.9129  -3.484 0.002003 **
## Row5             -9.4430     2.0804  -4.539 0.000147 ***
## Row6             -6.6172     1.9939  -3.319 0.002992 **
## Column2          -3.1909     2.2470  -1.420 0.169009
## Column3          -1.2547     2.8736  -0.437 0.666452
## Column4           0.5900     2.3487   0.251 0.803902
## Column5          -3.5812     2.1608  -1.657 0.111031
## Column6          -1.8595     3.2452  -0.573 0.572201
## Column7           2.7904     2.2779   1.225 0.232983
## Column8           0.3607     2.3256   0.155 0.878097
## Column9          -0.3207     2.4640  -0.130 0.897578
## Column10         -0.9091     2.3320  -0.390 0.700256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.859 on 23 degrees of freedom
##   (8 observations effacées parce que manquantes)
## Multiple R-squared:  0.817,  Adjusted R-squared:  0.6419
## F-statistic: 4.666 on 22 and 23 DF,  p-value: 0.0002542
##
## Analysis of Variance Table
##
## Response: DW_shoot_g
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Genotype   8 604.07  75.509  9.2408 1.276e-05 ***
## Row        5 142.31  28.462  3.4832   0.01726 *
## Column     9  92.41  10.268  1.2566   0.31131
## Residuals 23 187.94   8.171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "Summary for fixed effects model of FW_shoot_g"
##
## Call:
```

```
## lm(formula = fixed_formula, data = data)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -11.2013  -1.9550  -0.6745   2.2106  12.9658
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      88.7589     4.3487  20.410  < 2e-16 ***
## GenotypeEPPN1_H -10.9345     3.6966  -2.958 0.005882 **
## GenotypeEPPN1_L -22.3175     3.7303  -5.983 1.29e-06 ***
## GenotypeEPPN2_H -24.6194     3.8443  -6.404 3.91e-07 ***
## GenotypeEPPN2_L -22.9749     3.6644  -6.270 5.71e-07 ***
## GenotypeEPPN3_H -19.9662     3.3838  -5.900 1.63e-06 ***
## GenotypeEPPN3_L -16.9488     4.2173  -4.019 0.000347 ***
## GenotypeEPPN4_H -12.5686     3.7543  -3.348 0.002148 **
## GenotypeEPPN4_L -31.2305     3.6129  -8.644 9.21e-10 ***
## Row2              0.5827     2.8187   0.207 0.837569
## Row3             -2.7205     2.6893  -1.012 0.319548
## Row4              0.5547     2.8200   0.197 0.845339
## Row5             -5.0171     2.9625  -1.694 0.100380
## Row6             -3.7052     2.8220  -1.313 0.198824
## Column2           1.4920     3.6540   0.408 0.685848
## Column3           4.3469     4.2235   1.029 0.311338
## Column4           1.9934     4.0302   0.495 0.624353
## Column5           3.8870     3.6637   1.061 0.296908
## Column6           9.9256     4.4208   2.245 0.032028 *
## Column7           9.4191     3.7850   2.489 0.018407 *
## Column8           8.3263     3.6700   2.269 0.030396 *
## Column9           4.8191     4.0136   1.201 0.238964
## Column10          1.7144     3.6521   0.469 0.642048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.442 on 31 degrees of freedom
## Multiple R-squared:  0.8404, Adjusted R-squared:  0.7272
## F-statistic: 7.421 on 22 and 31 DF,  p-value: 3.668e-07
##
## Analysis of Variance Table
##
## Response: FW_shoot_g
##            Df Sum Sq Mean Sq F value   Pr(>F)
## Genotype    8 4027.9  503.49 17.0018 2.31e-09 ***
## Row         5  320.6   64.12  2.1652   0.0837 .
## Column      9  486.1   54.02  1.8240   0.1035
## Residuals  31  918.0   29.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "Summary for fixed effects model of DW_root_g"
##
## Call:
## lm(formula = fixed_formula, data = data)
##
## Residuals:
##     Min     1Q  Median      3Q     Max
## -6.1494 -1.7780  0.1214  1.8819  5.4548
```

```
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       20.2535     2.8977    6.990  1.1e-07 ***
## GenotypeEPPN1_H    0.6160     2.4957    0.247 0.806789
## GenotypeEPPN1_L   -4.9805     2.6191   -1.902 0.067201 .
## GenotypeEPPN2_H    3.0882     2.8002    1.103 0.279160
## GenotypeEPPN2_L   -1.3437     2.4487   -0.549 0.587378
## GenotypeEPPN3_H    0.4496     2.2571    0.199 0.843508
## GenotypeEPPN3_L    3.5893     2.9081    1.234 0.227019
## GenotypeEPPN4_H    3.1710     2.5257    1.256 0.219312
## GenotypeEPPN4_L   -9.1221     2.4174   -3.774 0.000737 ***
## Row2              -0.6827     1.9826   -0.344 0.733073
## Row3               1.4417     1.9474    0.740 0.465065
## Row4              -2.6523     1.9580   -1.355 0.186012
## Row5              -1.8923     2.0359   -0.929 0.360313
## Row6              -2.0299     1.9782   -1.026 0.313316
## Column2           -4.3234     2.7054   -1.598 0.120867
## Column3           -2.2682     2.8647   -0.792 0.434922
## Column4           -2.0407     2.6886   -0.759 0.453970
## Column5           -1.9043     2.4897   -0.765 0.450544
## Column6           -6.0325     3.3098   -1.823 0.078686 .
## Column7           -1.0016     2.5186   -0.398 0.693784
## Column8           -1.3183     2.4436   -0.540 0.593649
## Column9           -1.9554     2.6778   -0.730 0.471097
## Column10           2.2798     2.4347    0.936 0.356810
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.613 on 29 degrees of freedom
##   (2 observations effacées parce que manquantes)
## Multiple R-squared:  0.6673, Adjusted R-squared:  0.4149
## F-statistic: 2.644 on 22 and 29 DF,  p-value: 0.00749
##
## Analysis of Variance Table
##
## Response: DW_root_g
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Genotype   8 541.67  67.708  5.1871 0.0004448 ***
## Row        5  95.16  19.032  1.4580 0.2338305
## Column     9 122.37  13.596  1.0416 0.4324445
## Residuals 29 378.54  13.053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "Summary for fixed effects model of FW_root_g"
##
## Call:
## lm(formula = fixed_formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5089 -2.5356  0.5511  2.6074 10.1687
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       48.0814     4.8897    9.833 1.39e-10 ***
```

```
## GenotypeEPPN1_H  -2.3089      4.3966    -0.525  0.60361
## GenotypeEPPN1_L  -6.6037      4.2785    -1.543  0.13395
## GenotypeEPPN2_H  -2.8115      5.1646    -0.544  0.59049
## GenotypeEPPN2_L  -7.0269      4.2010    -1.673  0.10553
## GenotypeEPPN3_H   2.0322      4.0060     0.507  0.61591
## GenotypeEPPN3_L   6.7482      5.0648     1.332  0.19348
## GenotypeEPPN4_H   4.1184      4.3453     0.948  0.35135
## GenotypeEPPN4_L -13.2423      4.2934    -3.084  0.00455 **
## Row2             -3.0289      3.1572    -0.959  0.34558
## Row3              0.2770      3.2683     0.085  0.93306
## Row4             -2.7747      3.3189    -0.836  0.41022
## Row5             -3.8179      3.3468    -1.141  0.26363
## Row6             -7.5944      3.1647    -2.400  0.02330 *
## Column2           0.8536      4.0440     0.211  0.83435
## Column3          -0.4534      4.7408    -0.096  0.92448
## Column4           0.2597      4.4088     0.059  0.95344
## Column5           3.0998      4.2872     0.723  0.47566
## Column6          -6.9537      5.5286    -1.258  0.21887
## Column7           1.8336      4.1345     0.443  0.66083
## Column8           0.6228      4.0174     0.155  0.87792
## Column9          -0.6432      4.4657    -0.144  0.88650
## Column10          3.4837      4.0235     0.866  0.39393
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.939 on 28 degrees of freedom
##   (3 observations effacées parce que manquantes)
## Multiple R-squared:  0.6206, Adjusted R-squared:  0.3225
## F-statistic: 2.082 on 22 and 28 DF,  p-value: 0.03405
##
## Analysis of Variance Table
##
## Response: FW_root_g
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## Genotype    8 1126.67 140.834  3.9933 0.002895 **
## Row         5  320.87  64.175  1.8196 0.141365
## Column      9  167.72  18.636  0.5284 0.841371
## Residuals  28  987.50  35.268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "Summary for fixed effects model of Leaf_number"
##
## Call:
## lm(formula = fixed_formula, data = data)
##
## Residuals:
##     Min      1Q   Median      3Q      Max
## -0.79660 -0.26180 -0.02806  0.15571  1.18873
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.95820    0.52417  18.998 2.26e-16 ***
## GenotypeEPPN1_H -0.77385    0.43332  -1.786 0.086262 .
## GenotypeEPPN1_L -0.47112    0.46239  -1.019 0.318019
## GenotypeEPPN2_H -0.78416    0.47147  -1.663 0.108753
## GenotypeEPPN2_L -0.63077    0.43477  -1.451 0.159267
```

```
## GenotypeEPPN3_H -1.52126      0.37905   -4.013 0.000479 ***
## GenotypeEPPN3_L -0.93865      0.64640   -1.452 0.158904
## GenotypeEPPN4_H -0.59762      0.48744   -1.226 0.231604
## GenotypeEPPN4_L -1.00542      0.40772   -2.466 0.020876 *
## Row2            -0.09659      0.30236   -0.319 0.752041
## Row3            -0.01467      0.28483   -0.052 0.959336
## Row4             0.18895      0.31104    0.607 0.549013
## Row5            -0.13321      0.33062   -0.403 0.690443
## Row6             0.37379      0.34508    1.083 0.289063
## Column2         -0.09732      0.42224   -0.230 0.819600
## Column3         -0.29690      0.54270   -0.547 0.589174
## Column4         -0.48591      0.45013   -1.079 0.290678
## Column5         -0.19977      0.41474   -0.482 0.634224
## Column6         -0.07760      0.63853   -0.122 0.904245
## Column7          0.07498      0.41610    0.180 0.858449
## Column8         -0.13186      0.42131   -0.313 0.756899
## Column9         -0.10908      0.49235   -0.222 0.826468
## Column10        -0.67581      0.43464   -1.555 0.132551
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5558 on 25 degrees of freedom
##   (6 observations effacées parce que manquantes)
## Multiple R-squared:  0.5689, Adjusted R-squared:  0.1896
## F-statistic:   1.5 on 22 and 25 DF,  p-value: 0.1636
##
## Analysis of Variance Table
##
## Response: Leaf_number
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Genotype   8 6.7833 0.84792  2.7446 0.02535 *
## Row        5 1.5025 0.30051  0.9727 0.45347
## Column     9 1.9072 0.21191  0.6859 0.71461
## Residuals 25 7.7236 0.30894
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
for (trait in variables) {
  fit_models_random(endpoint_clean, trait)
}
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## [1] "Summary for random effects model of DW_shoot_g"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: random_formula
##    Data: data
##
## REML criterion at convergence: 207
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7270 -0.3655 -0.0142  0.2999  3.2050
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  Column   (Intercept) 2.107e-09 0.0000459
##  Row      (Intercept) 3.514e+00 1.8744523
##  Residual             8.913e+00 2.9854133
## Number of obs: 46, groups:  Column, 10; Row, 6
##
## Fixed effects:
##                Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      20.230      1.481  21.041  13.663 6.23e-12 ***
## GenotypeEPPN1_H  -4.392      1.915  34.258  -2.293  0.02808 *
## GenotypeEPPN1_L  -8.415      1.847  32.098  -4.555 7.16e-05 ***
## GenotypeEPPN2_H  -4.148      1.914  34.136  -2.167  0.03733 *
## GenotypeEPPN2_L  -5.972      1.915  34.258  -3.118  0.00367 **
## GenotypeEPPN3_H  -5.043      1.908  33.943  -2.643  0.01235 *
## GenotypeEPPN3_L  -7.914      1.995  33.028  -3.967  0.00037 ***
## GenotypeEPPN4_H  -5.801      1.832  31.716  -3.166  0.00340 **
## GenotypeEPPN4_L -13.258      1.784  32.711  -7.433 1.62e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) GEPPN1_H GEPPN1_L GEPPN2_H GEPPN2_L GEPPN3_H GEPPN3_L
## GntyEPPN1_H -0.595
## GntyEPPN1_L -0.586  0.463
## GntyEPPN2_H -0.580  0.493    0.450
## GntyEPPN2_L -0.595  0.514    0.463    0.493
## GntyEPPN3_H -0.594  0.484    0.482    0.481    0.484
## GntyEPPN3_L -0.533  0.433    0.420    0.461    0.433    0.443
## GntyEPPN4_H -0.581  0.469    0.464    0.447    0.469    0.460    0.417
## GntyEPPN4_L -0.617  0.491    0.502    0.487    0.491    0.516    0.449
##            GEPPN4_H
## GntyEPPN1_H
## GntyEPPN1_L
## GntyEPPN2_H
## GntyEPPN2_L
## GntyEPPN3_H
## GntyEPPN3_L
## GntyEPPN4_H
## GntyEPPN4_L  0.481
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
##
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## Genotype 582.29  72.786     8 32.322  8.1666 5.647e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## DW_shoot_g ~ Genotype + (1 | Row) + (1 | Column)
##               npar  logLik    AIC     LRT Df Pr(>Chisq)
## <none>          12 -103.48 230.96
## (1 | Row)       11 -104.87 231.75 2.7902  1    0.09484 .
## (1 | Column)    11 -103.48 228.96 0.0000  1    1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "Summary for random effects model of FW_shoot_g"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: random_formula
##    Data: data
##
## REML criterion at convergence: 304.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.52711 -0.42118 -0.05177  0.46252  2.52150
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept)  7.052   2.655
##  Row      (Intercept)  3.166   1.779
##  Residual             28.740   5.361
## Number of obs: 54, groups:  Column, 10; Row, 6
##
## Fixed effects:
##                 Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)       91.690      2.583 39.754  35.492  < 2e-16 ***
## GenotypeEPPN1_H  -11.952      3.350 41.337  -3.567 0.000929 ***
## GenotypeEPPN1_L  -23.119      3.358 40.946  -6.885 2.41e-08 ***
## GenotypeEPPN2_H  -25.708      3.397 42.685  -7.567 2.05e-09 ***
## GenotypeEPPN2_L  -22.969      3.332 40.700  -6.893 2.42e-08 ***
## GenotypeEPPN3_H  -19.912      3.213 36.884  -6.198 3.43e-07 ***
## GenotypeEPPN3_L  -15.357      3.482 42.468  -4.410 6.92e-05 ***
## GenotypeEPPN4_H  -12.045      3.348 41.290  -3.598 0.000850 ***
## GenotypeEPPN4_L  -30.688      3.313 40.279  -9.262 1.57e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) GEPPN1_H GEPPN1_L GEPPN2_H GEPPN2_L GEPPN3_H GEPPN3_L
## GntyEPPN1_H -0.654
## GntyEPPN1_L -0.639  0.503
## GntyEPPN2_H -0.643  0.544    0.503
## GntyEPPN2_L -0.644  0.531    0.511    0.494
## GntyEPPN3_H -0.627  0.486    0.485    0.487    0.487
## GntyEPPN3_L -0.627  0.482    0.444    0.490    0.467    0.450
## GntyEPPN4_H -0.650  0.504    0.513    0.489    0.501    0.484    0.542
## GntyEPPN4_L -0.631  0.518    0.487    0.483    0.510    0.515    0.440
##             GEPPN4_H
## GntyEPPN1_H
```

```
## GntyEPPN1_L
## GntyEPPN2_H
## GntyEPPN2_L
## GntyEPPN3_H
## GntyEPPN3_L
## GntyEPPN4_H
## GntyEPPN4_L  0.468
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## Genotype 3551.5  443.94      8 39.696  15.447 5.088e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## FW_shoot_g ~ Genotype + (1 | Row) + (1 | Column)
##              npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>         12 -152.33 328.65
## (1 | Row)      11 -152.83 327.65 1.0015  1     0.3169
## (1 | Column)   11 -153.50 329.00 2.3520  1     0.1251
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## [1] "Summary for random effects model of DW_root_g"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: random_formula
##    Data: data
##
## REML criterion at convergence: 250.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.92322 -0.79925  0.07467  0.69963  2.24034
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept)  0.0000  0.0000
##  Row      (Intercept)  0.8247  0.9081
##  Residual             13.0906  3.6181
## Number of obs: 52, groups:  Column, 10; Row, 6
##
## Fixed effects:
##                 Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)      17.7226     1.5432  38.7464  11.484 4.84e-14 ***
## GenotypeEPPN1_H  -0.4296     2.1185  41.7465  -0.203 0.840280
## GenotypeEPPN1_L  -2.5270     2.2088  40.5495  -1.144 0.259323
## GenotypeEPPN2_H   2.1976     2.2383  42.6922   0.982 0.331715
## GenotypeEPPN2_L  -0.9993     2.1413  42.9166  -0.467 0.643099
## GenotypeEPPN3_H   0.9982     2.1177  41.6077   0.471 0.639850
## GenotypeEPPN3_L   0.7584     2.1118  41.2289   0.359 0.721341
## GenotypeEPPN4_H   2.9464     2.1111  41.0863   1.396 0.170296
## GenotypeEPPN4_L  -8.7018     2.1177  41.6077  -4.109 0.000182 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) GEPPN1_H GEPPN1_L GEPPN2_H GEPPN2_L GEPPN3_H GEPPN3_L
## GntyEPPN1_H -0.691
## GntyEPPN1_L -0.658  0.479
## GntyEPPN2_H -0.655  0.485    0.452
## GntyEPPN2_L -0.694  0.516    0.480    0.487
## GntyEPPN3_H -0.691  0.503    0.487    0.480    0.505
## GntyEPPN3_L -0.689  0.502    0.481    0.483    0.507    0.505
## GntyEPPN4_H -0.689  0.505    0.481    0.472    0.510    0.502    0.504
## GntyEPPN4_L -0.691  0.503    0.487    0.480    0.505    0.513    0.505
##            GEPPN4_H
## GntyEPPN1_H
## GntyEPPN1_L
## GntyEPPN2_H
## GntyEPPN2_L
## GntyEPPN3_H
## GntyEPPN3_L
## GntyEPPN4_H
## GntyEPPN4_L  0.502
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
##
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##            Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## Genotype  548.3  68.538      8 40.751  5.2357 0.0001546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## DW_root_g ~ Genotype + (1 | Row) + (1 | Column)
##               npar  logLik    AIC     LRT Df Pr(>Chisq)
## <none>          12 -125.19 274.38
## (1 | Row)       11 -125.42 272.84 0.45809  1     0.4985
## (1 | Column)    11 -125.19 272.38 0.00000  1     1.0000
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## [1] "Summary for random effects model of FW_root_g"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: random_formula
##    Data: data
##
## REML criterion at convergence: 282.9
##
## Scaled residuals:
##      Min      1Q   Median      3Q      Max
## -1.58872 -0.72995 -0.09558  0.75594  1.84896
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept)  0.000   0.000
##  Row      (Intercept)  4.335   2.082
##  Residual             31.204   5.586
## Number of obs: 51, groups:  Column, 10; Row, 6
##
## Fixed effects:
##                Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      45.769      2.684  37.788  17.050  < 2e-16 ***
## GenotypeEPPN1_H  -1.872      3.437  39.122  -0.545  0.58905
## GenotypeEPPN1_L  -5.169      3.413  38.206  -1.514  0.13819
## GenotypeEPPN2_H  -1.395      3.655  40.860  -0.382  0.70477
## GenotypeEPPN2_L  -6.520      3.496  40.762  -1.865  0.06937 .
## GenotypeEPPN3_H   2.583      3.454  39.586   0.748  0.45899
## GenotypeEPPN3_L   2.565      3.606  39.591   0.711  0.48107
## GenotypeEPPN4_H   2.730      3.434  38.961   0.795  0.43140
## GenotypeEPPN4_L -12.033      3.454  39.586  -3.483  0.00122 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) GEPPN1_H GEPPN1_L GEPPN2_H GEPPN2_L GEPPN3_H GEPPN3_L
## GntyEPPN1_H -0.703
## GntyEPPN1_L -0.699  0.541
## GntyEPPN2_H -0.672  0.531    0.510
## GntyEPPN2_L -0.707  0.562    0.539    0.534
## GntyEPPN3_H -0.706  0.547    0.554    0.527    0.550
## GntyEPPN3_L -0.672  0.519    0.516    0.501    0.527    0.522
## GntyEPPN4_H -0.703  0.549    0.546    0.514    0.557    0.547    0.530
## GntyEPPN4_L -0.706  0.547    0.554    0.527    0.550    0.564    0.522
##            GEPPN4_H
## GntyEPPN1_H
## GntyEPPN1_L
## GntyEPPN2_H
## GntyEPPN2_L
## GntyEPPN3_H
## GntyEPPN3_L
## GntyEPPN4_H
## GntyEPPN4_L  0.547
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
##
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## Genotype 1146.9  143.36      8 39.159  4.5943 0.0005222 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## FW_root_g ~ Genotype + (1 | Row) + (1 | Column)
##               npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>          12 -141.44 306.88
## (1 | Row)       11 -142.13 306.27 1.3901  1     0.2384
## (1 | Column)    11 -141.44 304.88 0.0000  1     1.0000
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## [1] "Summary for random effects model of Leaf_number"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: random_formula
##    Data: data
##
## REML criterion at convergence: 76.7
##
## Scaled residuals:
##      Min      1Q   Median      3Q      Max
## -2.42764 -0.33747 -0.00033  0.34355  2.13002
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept) 0.008864 0.09415
##  Row      (Intercept) 0.000000 0.00000
##  Residual             0.277681 0.52695
## Number of obs: 48, groups:  Column, 10; Row, 6
##
## Fixed effects:
##                 Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)       9.8080     0.2406 36.2713  40.760  < 2e-16 ***
## GenotypeEPPN1_H  -0.8123     0.3237 38.9740  -2.509  0.01636 *
## GenotypeEPPN1_L  -0.4770     0.3262 36.0694  -1.462  0.15226
## GenotypeEPPN2_H  -0.8172     0.3381 38.9601  -2.417  0.02043 *
## GenotypeEPPN2_L  -0.6522     0.3228 38.8039  -2.020  0.05031 .
## GenotypeEPPN3_H  -1.4756     0.3204 35.2041  -4.606 5.18e-05 ***
## GenotypeEPPN3_L  -0.8081     0.3641 29.2696  -2.219  0.03439 *
## GenotypeEPPN4_H  -0.7984     0.3599 38.0912  -2.218  0.03256 *
## GenotypeEPPN4_L  -0.9759     0.3227 38.7008  -3.024  0.00441 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) GEPPN1_H GEPPN1_L GEPPN2_H GEPPN2_L GEPPN3_H GEPPN3_L
## GntyEPPN1_H -0.735
## GntyEPPN1_L -0.737  0.549
## GntyEPPN2_H -0.703  0.527    0.526
## GntyEPPN2_L -0.734  0.549    0.552    0.519
## GntyEPPN3_H -0.729  0.541    0.542    0.520    0.542
## GntyEPPN3_L -0.660  0.489    0.487    0.475    0.485    0.482
## GntyEPPN4_H -0.663  0.494    0.495    0.467    0.493    0.490    0.447
## GntyEPPN4_L -0.730  0.549    0.543    0.519    0.547    0.544    0.483
##            GEPPN4_H
## GntyEPPN1_H
## GntyEPPN1_L
## GntyEPPN2_H
## GntyEPPN2_L
## GntyEPPN3_H
## GntyEPPN3_L
## GntyEPPN4_H
## GntyEPPN4_L  0.488
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
##
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##          Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)
## Genotype 6.7708 0.84635     8 36.288  3.0479 0.009983 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## Leaf_number ~ Genotype + (1 | Row) + (1 | Column)
##              npar  logLik     AIC      LRT Df Pr(>Chisq)
## <none>         12 -38.336 100.672
## (1 | Row)      11 -38.336  98.672 0.000000  1     1.0000
## (1 | Column)   11 -38.368  98.736 0.064523  1     0.7995
```

2. Linear models with Plant_type

Model with X as Plant_type instead of Genotype, and row and column effects as random effects.
Plant_type is defined as H for Hybrid, L for pure Line and T for Tester.

```r
endpoint_clean$Plant_type <- as.factor(endpoint_clean$Plant_type)
endpoint_clean$Plant_type <- relevel(endpoint_clean$Plant_type, ref = "T") # T as base
level

fit_model_plant_type <- function(data, trait) {
  # Random effects model with Plant_type as a fixed effect
  model_formula <- as.formula(paste(trait, "~ Plant_type + (1|Row) + (1|Column)"))
  model <- lmer(model_formula, data)
  print(paste("Summary for mixed effects model of", trait))
  print(summary(model))
  print(anova(model))
  print(ranova(model))
}

for (trait in variables) {
  fit_model_plant_type(endpoint_clean, trait)
}
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## [1] "Summary for mixed effects model of DW_shoot_g"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: model_formula
##    Data: data
##
## REML criterion at convergence: 239.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8949 -0.4786  0.2091  0.4407  3.3828
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept)  0.00     0.000
##  Row      (Intercept)  0.00     0.000
##  Residual             12.69     3.562
## Number of obs: 46, groups:  Column, 10; Row, 6
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   20.750      1.454  43.000  14.268  < 2e-16 ***
## Plant_typeH   -5.845      1.658  43.000  -3.525  0.00102 **
## Plant_typeL   -9.845      1.658  43.000  -5.937 4.51e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Plnt_H
## Plant_typeH -0.877
## Plant_typeL -0.877  0.769
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
##
## Type III Analysis of Variance Table with Satterthwaite's method
##            Sum Sq Mean Sq NumDF DenDF F value   Pr(>F)
## Plant_type  481.1  240.55     2    43  18.957 1.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
## boundary (singular) fit: see help('isSingular')
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## DW_shoot_g ~ Plant_type + (1 | Row) + (1 | Column)
##              npar  logLik    AIC LRT Df Pr(>Chisq)
## <none>          6 -119.53 251.06
## (1 | Row)       5 -119.53 249.06   0  1          1
## (1 | Column)    5 -119.53 249.06   0  1          1
## [1] "Summary for mixed effects model of FW_shoot_g"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: model_formula
##    Data: data
##
## REML criterion at convergence: 362
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6718 -0.4739  0.1032  0.5149  2.3460
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept) 13.925   3.732
##  Row      (Intercept)  2.975   1.725
##  Residual             49.363   7.026
## Number of obs: 54, groups:  Column, 10; Row, 6
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   92.254      3.336  45.008  27.657  < 2e-16 ***
## Plant_typeH  -17.764      3.425  46.254  -5.187 4.63e-06 ***
## Plant_typeL  -23.693      3.421  46.146  -6.925 1.16e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Plnt_H
## Plant_typeH -0.821
## Plant_typeL -0.822  0.807
## Type III Analysis of Variance Table with Satterthwaite's method
##            Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## Plant_type 2389.6  1194.8     2 44.583  24.204 7.636e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## FW_shoot_g ~ Plant_type + (1 | Row) + (1 | Column)
##              npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>          6 -180.98 373.96
## (1 | Row)       5 -181.22 372.44 0.4772  1    0.48969
## (1 | Column)    5 -182.64 375.28 3.3144  1    0.06867 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## [1] "Summary for mixed effects model of DW_root_g"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: model_formula
##    Data: data
##
## REML criterion at convergence: 290.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.67931 -0.73966  0.05344  0.61744  2.29706
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept)  1.432   1.197
##  Row      (Intercept)  0.000   0.000
##  Residual             17.456   4.178
## Number of obs: 52, groups:  Column, 10; Row, 6
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   17.514      1.784  47.186   9.818 5.54e-13 ***
## Plant_typeH    1.776      1.963  47.894   0.904    0.370
## Plant_typeL   -2.709      1.962  47.639  -1.381    0.174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Plnt_H
## Plant_typeH -0.872
## Plant_typeL -0.872  0.791
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
##
## Type III Analysis of Variance Table with Satterthwaite's method
##            Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)
## Plant_type  219.3  109.65     2 48.066  6.2814 0.003771 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## DW_root_g ~ Plant_type + (1 | Row) + (1 | Column)
##              npar  logLik    AIC     LRT Df Pr(>Chisq)
## <none>          6 -145.16 302.33
## (1 | Row)       5 -145.16 300.33 0.00000  1     1.0000
## (1 | Column)    5 -145.40 300.79 0.46453  1     0.4955
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## [1] "Summary for mixed effects model of FW_root_g"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: model_formula
##    Data: data
##
## REML criterion at convergence: 325.5
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -2.0804 -0.7059  0.2179  0.7495  2.3252
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept)  0.0000  0.0000
##  Row      (Intercept)  0.2935  0.5417
##  Residual             43.5491  6.5992
## Number of obs: 51, groups:  Column, 10; Row, 6
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  45.8429     2.9649  45.7454  15.462   <2e-16 ***
## Plant_typeH   0.6121     3.2631  47.5309   0.188   0.8520
## Plant_typeL  -5.7730     3.2630  47.4713  -1.769   0.0833 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Plnt_H
## Plant_typeH -0.904
## Plant_typeL -0.904  0.822
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
##
## Type III Analysis of Variance Table with Satterthwaite's method
##            Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)
## Plant_type 498.25  249.12     2 46.096  5.7205 0.006037 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## FW_root_g ~ Plant_type + (1 | Row) + (1 | Column)
##             npar  logLik    AIC       LRT Df Pr(>Chisq)
## <none>         6 -162.76 337.52
## (1 | Row)      5 -162.76 335.52 0.0061995  1     0.9372
## (1 | Column)   5 -162.76 335.52 0.0000000  1     1.0000
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## [1] "Summary for mixed effects model of Leaf_number"
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: model_formula
##    Data: data
##
## REML criterion at convergence: 82.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9865 -0.2648  0.2498  0.3663  1.7729
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Column   (Intercept) 0.01502  0.1226
##  Row      (Intercept) 0.00000  0.0000
##  Residual             0.29420  0.5424
## Number of obs: 48, groups:  Column, 10; Row, 6
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   9.8352     0.2503 42.6609  39.294  < 2e-16 ***
## Plant_typeH  -1.0282     0.2745 44.3488  -3.746 0.000516 ***
## Plant_typeL  -0.7520     0.2753 44.9498  -2.732 0.008974 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Plnt_H
## Plant_typeH -0.889
## Plant_typeL -0.894  0.812
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
##
## Type III Analysis of Variance Table with Satterthwaite's method
##            Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)
## Plant_type 4.2103  2.1051     2 44.289  7.1555 0.002029 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## Leaf_number ~ Plant_type + (1 | Row) + (1 | Column)
##              npar  logLik    AIC     LRT Df Pr(>Chisq)
## <none>          6 -41.137 94.275
## (1 | Row)       5 -41.137 92.275 0.00000  1     1.0000
## (1 | Column)    5 -41.221 92.443 0.16815  1     0.6818
```

3. Linear models with asreml library

```
modasreml <- asreml(fixed = DW_shoot_g ~ Genotype,
                    random = ~ Row + Column,
                    residual = ~  NULL,
                    data = endpoint_clean)
```

```
## ASReml Version 4.2 09/06/2024 16:49:15
##           LogLik      Sigma2   DF     wall
## 1      -70.44230   9.679677   37   16:49:15
## 2      -69.88044   9.722756   37   16:49:15  (  1 restrained)
## 3      -69.52554   9.209241   37   16:49:15  (  1 restrained)
## 4      -69.48072   8.983979   37   16:49:15  (  1 restrained)
## 5      -69.47788   8.924269   37   16:49:15  (  1 restrained)
## 6      -69.47781   8.914247   37   16:49:15
```

```
plot(modasreml)
```
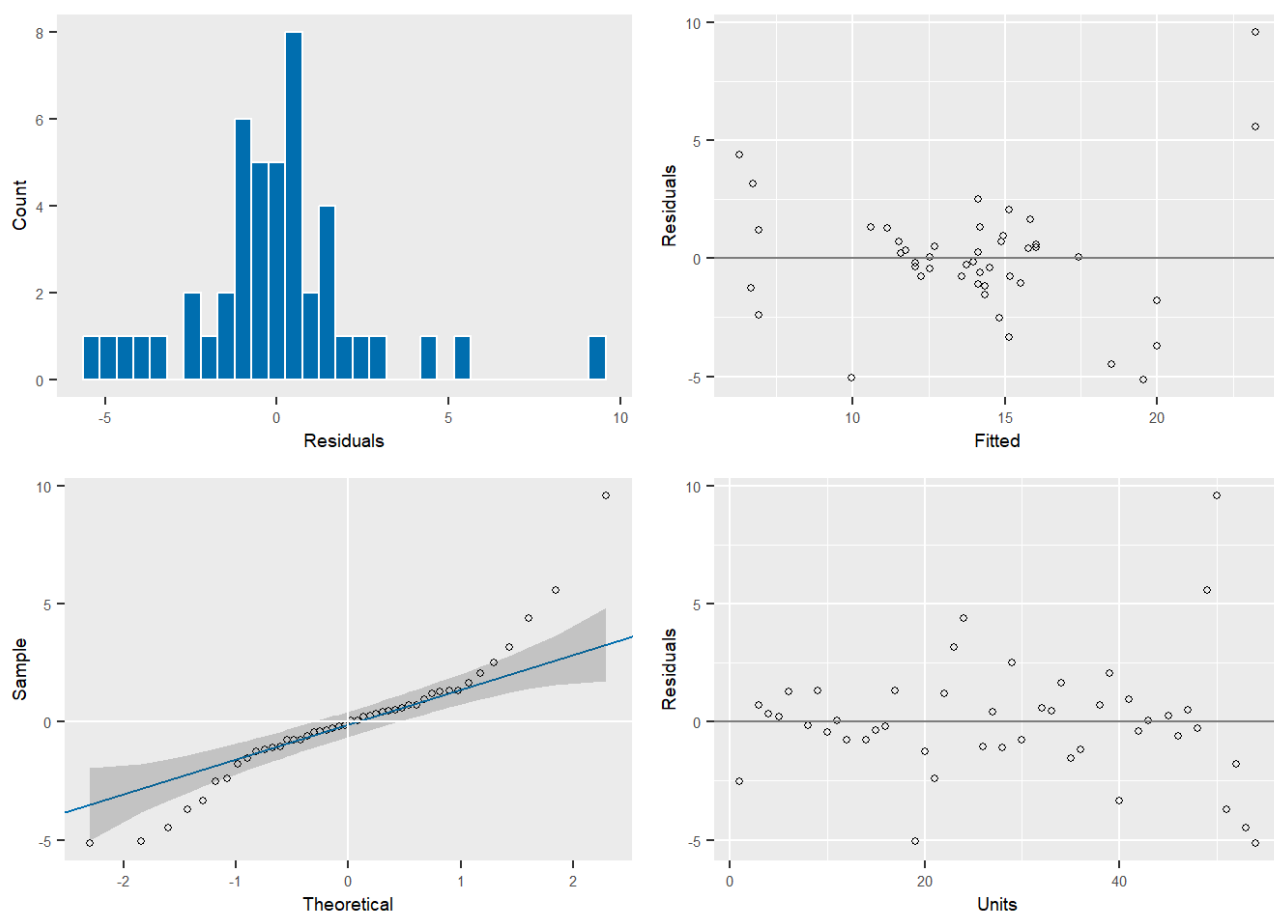


```
summary(modasreml)$varcomp
```

```
##              component   std.error   z.ratio  bound  %ch
## Row       3.513371e+00   3.285813   1.069255      P  0.2
## Column    3.142980e-06         NA        NA      B   NA
## units!R   8.914247e+00   2.247272   3.966696      P  0.0
```

4. Linear models with Soil variable

Model with Soil as explicative variable.

5. Linear models with Soil variable with Plant_type