17. R Analysis of Split-Plot Experiments

Field Split-Plot Experiment

Field

Block 1	Genotype C				Genotype A				Genotype B			
	0	100	150	50	50	100	150	0	150	100	50	0
Block 2	Genotype B				Genotype A				Genotype C			
	150	100	50	0	0	50	150	100	100	50	150	0
Block 3	Genotype A				Genotype B				Genotype C			
	100	50	0	150	0	100	150	50	50	100	150	0
Block 4	Genotype B				Genotype C				Genotype A			
	0	50	100	150	150	100	50	0	50	150	100	0

Install and Load Packages for Linear Mixed-Model Analysis in R

PACKAGE FOR FITTING LINEAR MIXED-EFFECT MODELS WITH lmer FUNCTION
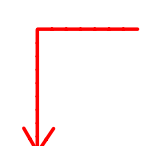
```
> install.packages("lme4")

> install.packages("lmerTest")

> library(lme4)

> library(lmerTest)
```

PACKAGE FOR GETTING CORRECT p-VALUES WHEN PERFORMING LINEAR MIXED-EFFECT MODEL ANALYSIS IN R

Read Data

```
> fd = read.delim(
  "https://dnett.github.io/S510/FieldSplitPlotData.txt")

> head(fd)
  block geno fert     y
1     1    1    0 148.7
2     1    1   50 150.4
3     1    1  100 166.7
4     1    1  150 156.5
5     1    2    0 162.5
6     1    2   50 168.6
```

```
> #Define factors and shorten names.

> y = fd$y

> b = factor(fd$block)

> g = factor(fd$geno)

> f = factor(fd$fert/50+1)
```

Fit the Model

```
> o = lmer(y ~ g + f + g:f + (1 | b) + (1 | b:g))
```

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + b_k + w_{ik} + e_{ijk}$$

$$b_k \overset{iid}{\sim} N(0, \sigma_b^2)$$

$$w_{ik} \overset{iid}{\sim} N(0, \sigma_w^2)$$

$$e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2)$$

```
> summary(o)

Linear mixed model fit by REML.
```

PROVIDED WHEN lmer Test PACKAGE IS USED

```
t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: y ~ g + f + g:f + (1 | b) + (1 | b:g)

REML criterion at convergence: 275.1
```
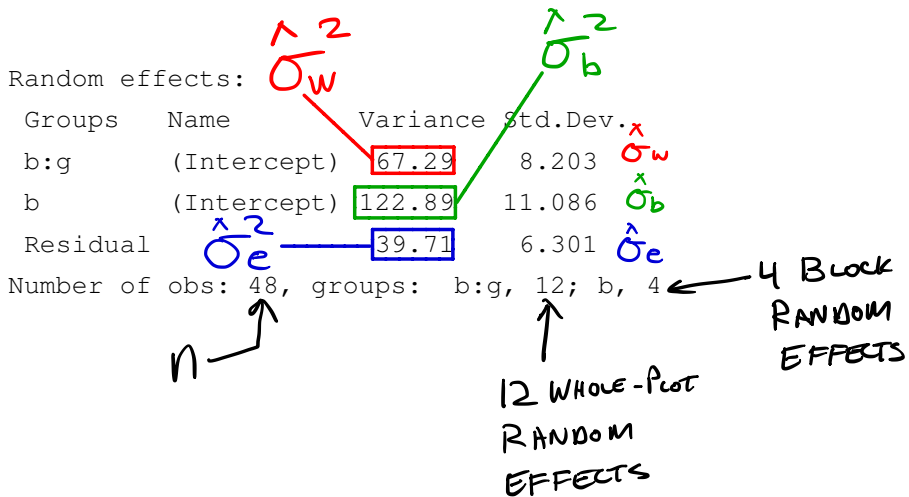
WE WILL LEARN ABOUT THIS LATER (SLIDE SET 20)

```
Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.43560 -0.60942 -0.07892  0.53263  2.09822
```

IGNORE FOR NOW

$\hat{\sigma}_w^2$ $\hat{\sigma}_b^2$

```
Random effects:
 Groups    Name          Variance  Std.Dev.
 b:g       (Intercept)    67.29     8.203      σ̂w
 b         (Intercept)   122.89    11.086      σ̂b
 Residual                 39.71     6.301      σ̂e
Number of obs: 48, groups:  b:g, 12; b, 4
```

$\hat{\sigma}_e^2$

n

4 BLOCK RANDOM EFFECTS

12 WHOLE-PLOT RANDOM EFFECTS

$\widehat{\underset{\sim}{\beta}}$

SATTERTHWAITE WHEN NEEDED

```
Fixed effects:

              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)    126.025      7.581    6.740  16.624 1.01e-06 ***
g2              22.500      7.314   11.154   3.076 0.010384 *
g3             -12.275      7.314   11.154  -1.678 0.121064
f2              13.750      4.456   27.000   3.086 0.004649 **
f3              22.425      4.456   27.000   5.033 2.79e-05 ***
f4              19.025      4.456   27.000   4.270 0.000216 ***
g2:f2           -2.825      6.301   27.000  -0.448 0.657499
g3:f2           17.875      6.301   27.000   2.837 0.008540 **
g2:f3            3.750      6.301   27.000   0.595 0.556721
g3:f3           24.525      6.301   27.000   3.892 0.000588 ***
g2:f4            8.600      6.301   27.000   1.365 0.183583
g3:f4           35.800      6.301   27.000   5.681 4.92e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Correlation of Fixed Effects:
     (Intr) g2     g3     f2     f3     f4     g2:f2  g3:f2  g2:f3  g3:f3  g2:f4
g2   -0.482
g3   -0.482  0.500
f2   -0.294  0.305  0.305
f3   -0.294  0.305  0.305  0.500
f4   -0.294  0.305  0.305  0.500  0.500
g2:f2  0.208 -0.431 -0.215 -0.707 -0.354 -0.354
g3:f2  0.208 -0.215 -0.431 -0.707 -0.354 -0.354  0.500
g2:f3  0.208 -0.431 -0.215 -0.354 -0.707 -0.354  0.500  0.250
g3:f3  0.208 -0.215 -0.431 -0.354 -0.707 -0.354  0.250  0.500  0.500
g2:f4  0.208 -0.431 -0.215 -0.354 -0.354 -0.707  0.500  0.250  0.500  0.250
g3:f4  0.208 -0.215 -0.431 -0.354 -0.354 -0.707  0.250  0.500  0.250  0.500  0.500
```

lmerTest is not perfect.
The DF should be 6, 27, 27 here.

This difference is minor and the p-values are not affected in an important way

```
> anova(o)
Type III Analysis of Variance Table with Satterthwaite's method
      Sum Sq Mean Sq NumDF   DenDF F value    Pr(>F)
g      673.3  336.65     2  6.0013  8.4784 0.0178463 *
f     8737.7 2912.57     3 26.9999 73.3529 4.234e-13 ***
g:f   1557.3  259.56     6 26.9999  6.5369 0.0002381 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

lmerTest Provides An ls_means Function To Get LSMEANS As In SAS

```
> ls_means(o)
Least Squares Means table:

     Estimate Std. Error  df t value   lower   upper   Pr(>|t|)
g1    139.825      7.073 5.1  19.769 121.779 157.870 4.870e-06 ***
g2    164.706      7.073 5.1  23.287 146.661 182.752 2.120e-06 ***
g3    147.100      7.073 5.1  20.797 129.054 165.145 3.765e-06 ***
f1    129.433      6.296 3.4  20.558 110.696 148.171  0.000107 ***
f2    148.200      6.296 3.4  23.539 129.463 166.937 6.758e-05 ***
f3    161.283      6.296 3.4  25.617 142.546 180.020 5.071e-05 ***
f4    163.258      6.296 3.4  25.930 144.521 181.995 4.865e-05 ***
```

```
        Estimate Std. Error  df t value   lower    upper  Pr(>|t|)
g1:f1    126.025      7.581 6.7  16.624 107.958  144.092  1.012e-06 ***
g2:f1    148.525      7.581 6.7  19.592 130.458  166.592  3.411e-07 ***
g3:f1    113.750      7.581 6.7  15.005  95.683  131.817  1.986e-06 ***
g1:f2    139.775      7.581 6.7  18.437 121.708  157.842  5.102e-07 ***
g2:f2    159.450      7.581 6.7  21.033 141.383  177.517  2.129e-07 ***
g3:f2    145.375      7.581 6.7  19.176 127.308  163.442  3.932e-07 ***
g1:f3    148.450      7.581 6.7  19.582 130.383  166.517  3.423e-07 ***
g2:f3    174.700      7.581 6.7  23.044 156.633  192.767  1.159e-07 ***
g3:f3    160.700      7.581 6.7  21.198 142.633  178.767  2.021e-07 ***
g1:f4    145.050      7.581 6.7  19.133 126.983  163.117  3.991e-07 ***
g2:f4    176.150      7.581 6.7  23.236 158.083  194.217  1.097e-07 ***
g3:f4    168.575      7.581 6.7  22.236 150.508  186.642  1.470e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

  Confidence level: 95%
  Degrees of freedom method: Satterthwaite
```

```
> betahat=fixef(o)
> betahat
(Intercept)            g2            g3
    126.025        22.500       -12.275

     f2            f3            f4
 13.750        22.425        19.025

  g2:f2      g3:f2      g2:f3      g3:f3      g2:f4      g3:f4
 -2.825     17.875      3.750     24.525      8.600     35.800
```

```
> #The following table shows the cell means
> #in terms of the R parameterization.
> ###########################################################
> #
> #    f1=0units  f2=50units      f3=100units      f4=150units
> #
> #g1 | mu       | mu    +f2       | mu    +f3       | mu    +f4       |
> #   |          |                 |                 |                 |
> #g2 | mu+g2    | mu+g2+f2+g2f2   | mu+g2+f3+g2f3   | mu+g2+f4+g2f4   |
> #   |          |                 |                 |                 |
> #g3 | mu+g3    | mu+g3+f2+g3f2   | mu+g3+f3+g3f3   | mu+g3+f4+g3f4   |
> #
> ###########################################################
```

AVERAGE OF ROW 1 FROM SLIDE 15:

$$\frac{(\mu) + (\mu + f2) + (\mu + f3) + (\mu + f4)}{4} = \mu + \frac{f2 + f3 + f4}{4}$$

```
> #Coefficients for geno 1 marginal mean
>
> C1 = matrix(c(1,
+                0, 0,
+                1/4, 1/4, 1/4,
+                0, 0, 0, 0, 0, 0), nrow=1)
>
```

$$1 \cdot \mu + 0 \cdot g2 + 0 \cdot g3 + \tfrac{1}{4} \cdot f2 + \tfrac{1}{4} f3 + \tfrac{1}{4} f4$$

$$= \mu + \frac{f2 + f3 + f4}{4}$$

```
> #Coefficients for geno 1 - geno 2 marginal mean
>
> C2 = matrix(c(0,
+             -1, 0,
+              0, 0, 0,
+             -1/4, 0, -1/4, 0, -1/4, 0), nrow=1)
>
>
```

```
> #Coefficients for geno 1 - geno 2 with no fertilizer
>
> C3 = matrix(c(0,
+              -1, 0,
+               0, 0, 0,
+               0, 0, 0, 0, 0, 0), nrow=1)
>
```

Another lmerTest Function

WITH joint SET TO FALSE (F), WE GET A
SEPARATE TEST FOR EACH ROW OF
L = C RATHER THAN ONE F TEST OF $C\beta = 0$.

```
> C = rbind(C1, C2, C3)
> contest(o, L = C, joint = F, confint = T)
    Estimate Std. Error    df   t value      lower       upper       Pr(>|t|)
1  139.82500   7.072992  5.129 19.768862  121.77948  157.870523  4.870369e-06
2  -24.88125   6.213453  6.001 -4.004416  -40.08425   -9.678254  7.080208e-03
3  -22.50000   7.314154 11.154 -3.076227  -38.57124   -6.428757  1.038405e-02
```

```
> #The degrees of freedom, sums of squares,
> #and mean squares from a sequential ANOVA
> #table could be used to estimate variance
> #components and compute test statistics.
>
> a = anova(lm(y ~ b + g + b:g + f + g:f))
```

```
> a
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
b          3 5349.5 1783.16 44.9089 1.252e-10 ***
g          2 5237.2 2618.62 65.9500 4.057e-11 ***
f          3 8737.7 2912.57 73.3531 4.233e-13 ***
b:g        6 1853.4  308.90  7.7796 6.355e-05 ***
g:f        6 1557.3  259.56  6.5370 0.0002381 ***
Residuals 27 1072.1   39.71
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> #For example, based on expected mean squares
> #presented in slide set 15, an unbiased
> #estimator of variance for the whole-plot
> #random effects is (MSbg - MSe) / 4.
>
> MSbg = a[4,3]
> MSbg
[1] 308.8985
> MSe = a[6,3]
> MSe
[1] 39.70613
> (MSbg - MSe) / 4
[1] 67.2981
```

```
> #The F tests and p-values in the
> #ANOVA table "a" all use MSe as the
> #denominator becuase the lm function
> #implies a Gauss-Markov model rather
> #than a linear mixed-effects model.
> #This is fine for f and g:f, but not
> #for the other lines of the ANOVA
> #table.
```

```
> #The correct F statistic for testing
> #for genotype main effects is
> #MSg/MSbg
>
> MSg = a[2,3]
> MSg
[1] 2618.619
>
> MSbg = a[4,3]
> MSbg
[1] 308.8985
>
> MSg / MSbg
[1] 8.47728   ≈ 8.4784  GIVEN AS F STAT ON SLIDE 11
```

```
> #When the experiment is balanced, whole-plot-factor analysis
> #can be accomplished by computing the average for each
> #whole-plot experimental unit and then analyzing those
> #averages.
>
> d = aggregate(y, by = list(b, g), FUN = mean)
> names(d) = c("block", "geno", "wpaverage")
```

```
> d
   block geno wpaverage
1      1    1   155.575
2      2    1   134.075
3      3    1   135.100
4      4    1   134.550
5      1    2   173.100
6      2    2   162.925
7      3    2   154.600
8      4    2   168.200
9      1    3   177.250
10     2    3   142.800
11     3    3   137.325
12     4    3   131.025
```

```
> anova(lm(wpaverage ~ block + geno, data = d))
Analysis of Variance Table

Response: wpaverage
          Df  Sum Sq Mean Sq F value  Pr(>F)
block      3 1337.37  445.79  5.7726 0.03346 *
geno       2 1309.31  654.65  8.4773 0.01786 *
Residuals  6  463.35   77.22
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```
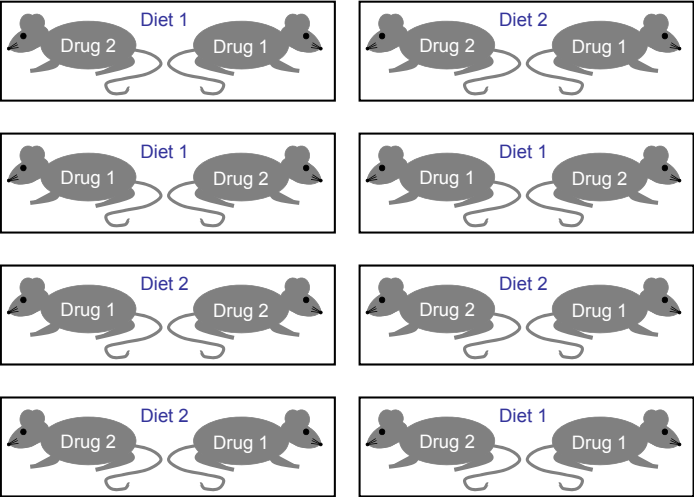
RESULT VERY SIMILAR TO SLIDE 11

```
> #Now analyze the diet drug split-plot experiment.
> #The main difference here is that the whole-plot
> #part of the experiment is a CRD rather than an
> #RCBD.
```

Diet and Drug Split-Plot Experiment

```
> ddd = read.delim(
"https://dnett.github.io/S510/DietDrugSplitPlotData.txt")

> head(ddd)
  litter diet drug    y
1      1    1    2 18.8
2      1    1    1 10.3
3      2    2    2 14.5
4      2    2    1 18.5
5      3    1    1 16.8
6      3    1    2 27.2
```

```
> ddd$litter = factor(ddd$litter)
> ddd$diet = factor(ddd$diet)
> ddd$drug = factor(ddd$drug)
>
> o = lmer(y ~ diet + drug + diet:drug + (1 | litter), data=ddd)
```

$$
\begin{aligned}
y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \ell_{ik} + e_{ijk} \\
\ell_{ik} &\overset{iid}{\sim} N(0, \sigma_\ell^2) \\
e_{ijk} &\overset{iid}{\sim} N(0, \sigma_e^2)
\end{aligned}
$$

```
> summary(o)
Linear mixed model fit by REML.

t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: y ~ diet + drug + diet:drug + (1 | litter)
   Data: ddd

REML criterion at convergence: 57.4

Scaled residuals:
     Min       1Q   Median       3Q      Max
-0.95525 -0.51772  0.05178  0.51149  1.02906
```

```
Random effects:
 Groups   Name        Variance Std.Dev.
 litter   (Intercept) 9.7871   3.128
 Residual             0.9408   0.970
Number of obs: 16, groups:  litter, 8
```

```
Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  13.5750     1.6377  6.5492   8.289 0.000105 ***
diet2         2.9250     2.3160  6.5492   1.263 0.249713
drug2         9.7500     0.6859  6.0000  14.216 7.57e-06 ***
diet2:drug2 -11.8500     0.9700  6.0000 -12.217 1.83e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Correlation of Fixed Effects:
           (Intr) diet2  drug2
diet2      -0.707
drug2      -0.209  0.148
diet2:drug2 0.148 -0.209 -0.707
```

```
> anova(o)
Type III Analysis of Variance Table with Satterthwaite's method
          Sum Sq Mean Sq NumDF DenDF  F value      Pr(>F)
diet       1.651   1.651     1     6   1.7548 0.2334898
drug      58.522  58.522     1     6  62.2028 0.0002202 ***
diet:drug 140.422 140.422    1     6 149.2533 1.831e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> ls_means(o)
Least Squares Means table:

            Estimate Std. Error  df t value   lower   upper  Pr(>|t|)
diet1        18.4500     1.6014  6.0 11.5214 14.5316 22.3684 2.569e-05 ***
diet2        15.4500     1.6014  6.0  9.6480 11.5316 19.3684 7.102e-05 ***
drug1        15.0375     1.1580  6.5 12.9856 12.2606 17.8144 6.455e-06 ***
drug2        18.8625     1.1580  6.5 16.2887 16.0856 21.6394 1.522e-06 ***
diet1:drug1  13.5750     1.6377  6.5  8.2892  9.6478 17.5022 0.0001047 ***
diet2:drug1  16.5000     1.6377  6.5 10.0753 12.5728 20.4272 3.169e-05 ***
diet1:drug2  23.3250     1.6377  6.5 14.2428 19.3978 27.2522 3.589e-06 ***
diet2:drug2  14.4000     1.6377  6.5  8.7930 10.4728 18.3272 7.320e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

  Confidence level: 95%
  Degrees of freedom method: Satterthwaite
```

```
> #The following table shows the cell means
> #in terms of the R parameterization.
> ############################################################
> #
> #            drug 1            drug 2
> #
> #diet 1     mu                mu + drug2
> #
> #diet 2     mu + diet2        mu + diet2 + drug2 + diet2:drug2
> #
> ############################################################
```

```
> #Coefficients for diet 1 - diet 2 marginal mean
>
> C1 = matrix(c(0, -1, 0, -1/2), nrow = 1)
>
```

```
> #Coefficients for drug 1 - drug 2 marginal mean
>
> C2 = matrix(c(0, 0, -1, -1/2), nrow = 1)
>
```

```
> #Coefficients for diet 1 - diet 2 for drug 2
>
> C3 = matrix(c(0, -1, 0, -1), nrow=1)
>
```

```
> #Coefficients for drug 1 - drug 2 for diet 2
>
> C4 = matrix(c(0, 0, -1, -1), nrow=1)
>
```

```
> C = rbind(C1, C2, C3, C4)
> contest(o, L = C, joint = F, confint = T)
  Estimate Std. Error       df    t value      lower      upper Pr(>|t|)
1    3.000  2.2646744 6.000000  1.324694 -2.5414586  8.541459  0.23349
2   -3.825  0.4849828 6.000000 -7.886877 -5.0117102 -2.638290  0.00022
3    8.925  2.3160221 6.549174  3.853590  3.3711210 14.478879  0.00712
4    2.100  0.6858693 6.000000  3.061808  0.4217383  3.778262  0.02217
```

```
> #The F statistic for diet main effects
> #is also given by the ratio
> #MSdiet / MSlitter from the appropriate
> #sequential ANOVA table. This can be
> #shown formally, but MSlitter makes
> #sense as the denominator mean square
> #because litters are the experimental
> #units for the factor diet.
```

```
> a = anova(
+     lm(y ~ diet + litter + drug + diet:drug,
+        data=ddd))
> a
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
diet       1  36.000  36.000  38.264 0.0008214 ***
litter     6 123.090  20.515  21.805 0.0007887 ***
drug       1  58.523  58.523  62.203 0.0002202 ***
diet:drug  1 140.423 140.423 149.253 1.831e-05 ***
Residuals  6   5.645   0.941
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> MSdiet = a[1,3]
> dfdiet = a[1,1]
> MSlitter = a[2,3]
> dflitter = a[2,1]
> F = MSdiet / MSlitter
> p = 1 - pf(F, dfdiet, dflitter)
> F
[1] 1.754814
> p
[1] 0.2334898
```

```
> #When the experiment is balanced, whole-plot-factor analysis
> #can be accomplished by computing the average for each
> #whole-plot experimental unit and then analyzing those
> #averages.
```

```
> d = aggregate(ddd$y, by = list(ddd$diet, ddd$litter),
                 FUN = mean)

> names(d) = c("diet", "litter", "wpaverage")
```

```
> d
  diet litter wpaverage
1    1      1     14.55
2    2      2     16.50
3    1      3     22.00
4    1      4     16.00
5    2      5     11.60
6    2      6     17.00
7    2      7     16.70
8    1      8     21.25
```

```
> anova(lm(wpaverage ~ diet, data = d))
Analysis of Variance Table

Response: wpaverage
          Df Sum Sq Mean Sq F value Pr(>F)
diet       1 18.000  18.000  1.7548 0.2335
Residuals  6 61.545  10.258
```