

Split plot analyses

Philip Dixon

11/14/2022

Split plot analyses in R

Really, the only new thing this week is the combination of two old things: specifying random effects (week 3) and specifying factorial treatment structures (weeks 6 and 7). The computer handles all the ugly details. I illustrate this using the PA study.

These data are simulated based on characteristics of a real study.

Reminder about the study design: Schools grouped into blocks, each with 2 similar schools. Treatments randomly assigned to schools within block: either intervention or control. PA measured on kids within schools. Second “treatment” factor is gender: boy or girl.

There are two data sets: pa2.xlsx has the balanced subset. This has 5 boys and 5 girls in each school. pa.csv has the full data set. Unequal numbers of boys and girls in each school. These data are similar to, but identical with, the data used as the class example.

```
library(emmeans)
library(lme4)
```

Let’s look at the balanced dataset first:

```
## Loading required package: Matrix
```

```
library(readxl)
```

```
pa2 <- read_excel('../data/pa2.xlsx')
pa2$block.f <- factor(pa2$block)
pa2$school.f <- factor(pa2$school)
pa2$gender.f <- factor(pa2$gender)
```

The bs variable identifies each school (combination of block and treatment). Let’s check that the data are as advertised.

```
with(pa2, table(bs, gender))
```

```
##           gender
## bs      Female Male
## 1/Active      5    5
## 1/Contro      5    5
## 10/Activ      5    5
## 10/Contr      5    5
## 2/Active      5    5
## 2/Contro      5    5
## 3/Active      5    5
## 3/Contro      5    5
## 4/Active      5    5
```

```
## 4/Contro      5    5
## 5/Active      5    5
## 5/Contro     5    5
## 6/Active      5    5
## 6/Contro     5    5
## 7/Active      5    5
## 7/Contro     5    5
## 8/Active      5    5
## 8/Contro     5    5
## 9/Active      5    5
## 9/Contro     5    5
```

Yes: 5 girls and 5 boys in each combination of block and treatment.

Fit a split plot model.

```
pa2.lme <- lmer(logmvpa ~ block.f + school.f +
  gender.f + gender.f:school.f +
  (1| block.f:school.f), data=pa2)
summary(pa2.lme)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: logmvpa ~ block.f + school.f + gender.f + gender.f:school.f +
## (1 | block.f:school.f)
## Data: pa2
##
## REML criterion at convergence: 398.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.68614 -0.69111  0.04575  0.71437  2.93629
##
## Random effects:
## Groups          Name          Variance Std.Dev.
## block.f:school.f (Intercept) 0.03039  0.1743
## Residual                0.38600  0.6213
## Number of obs: 200, groups: block.f:school.f, 20
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      1.9582    0.2045   9.577
## block.f2         -0.6395    0.2627  -2.435
## block.f3         -0.4990    0.2627  -1.900
## block.f4         -1.0155    0.2627  -3.866
## block.f5          0.5265    0.2627   2.004
## block.f6          0.2350    0.2627   0.895
## block.f7         -0.4080    0.2627  -1.553
## block.f8          0.1020    0.2627   0.388
## block.f9          0.0085    0.2627   0.032
## block.f10        -0.2900    0.2627  -1.104
## school.fContro   -0.2666    0.1467  -1.817
## gender.fMale      0.1696    0.1243   1.365
## school.fContro:gender.fMale -0.1948    0.1757  -1.109
##
## Correlation matrix not shown by default, as p = 13 > 12.
```

```
## Use print(x, correlation=TRUE) or
##   vcov(x)           if you need it
```

This model has blocks, a 2 way factorial treatment structure and a random effect for the main plot error (block:school).

First thing to check: do we have the correct numbers of levels for each random effect?

Observations: 200, check

Schools: 20, check

We'll use emmeans for both the ANOVA and the "after the ANOVA" analyses. Here's the type III ANOVA table.

```
pa2.emm <- emmeans(pa2.lme, c('school.f', 'gender.f'))
joint_tests(pa2.emm)
```

```
## model term      df1 df2 F.ratio p.value
## school.f        1   9   9.602  0.0127
## gender.f         1 178   0.675  0.4123
## school.f:gender.f 1 178   1.229  0.2691
```

How big is the difference between intervention and control?

```
pa2.school <- emmeans(pa2.lme, 'school.f')
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
pa2.school
```

```
## school.f emmean    SE df lower.CL upper.CL
## Active    1.84 0.0831  9    1.66    2.03
## Contro    1.48 0.0831  9    1.29    1.67
##
## Results are averaged over the levels of: block.f, gender.f
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

```
summary(contrast(pa2.school, list(diff=c(1, -1))),
        infer=c(T,T))
```

```
## contrast estimate    SE df lower.CL upper.CL t.ratio p.value
## diff              0.364 0.117  9    0.0983    0.63  3.099  0.0127
##
## Results are averaged over the levels of: block.f, gender.f
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

What about pairwise comparisons among treatment/gender groups? We already have those emmeans stored in pa2.emm.

```
pa2.emm
```

```
## school.f gender.f emmean    SE    df lower.CL upper.CL
## Active  Female    1.76 0.104 21.6    1.54    1.98
## Contro  Female    1.49 0.104 21.6    1.28    1.71
## Active  Male      1.93 0.104 21.6    1.71    2.15
## Contro  Male      1.47 0.104 21.6    1.25    1.68
##
## Results are averaged over the levels of: block.f
```

```
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

```
pairs(pa2.emm)
```

```
## contrast estimate SE df t.ratio p.value
## Active Female - Contro Female 0.2666 0.147 21.6 1.817 0.2927
## Active Female - Active Male -0.1696 0.124 178.0 -1.365 0.5230
## Active Female - Contro Male 0.2918 0.147 21.6 1.989 0.2229
## Contro Female - Active Male -0.4362 0.147 21.6 -2.974 0.0333
## Contro Female - Contro Male 0.0252 0.124 178.0 0.203 0.9970
## Active Male - Contro Male 0.4614 0.147 21.6 3.145 0.0229
```

```
##
```

```
## Results are averaged over the levels of: block.f
```

```
## Degrees-of-freedom method: kenward-roger
```

```
## P value adjustment: tukey method for comparing a family of 4 estimates
```

You should notice the 2 standard errors and 2 df in this table. Which depends on whether the comparison is within a school (i.e. within a mini-block) or between schools. Comparisons between schools have a larger se and smaller df.

```
pa <- read.csv('../data/pa.csv', as.is=T)
pa$block.f <- factor(pa$block)
pa$school.f <- factor(pa$school)
pa$gender.f <- factor(pa$gender)
```

Now let's look at the full (unbalanced data set) The bs variable identifies each school (combination of block and treatment). Let's check that the data are as advertised.

```
with(pa, table(bs, gender))
```

```
##          gender
## bs      Female Male
## 1/Active    19   14
## 1/Contro    13   20
## 10/Active   16   16
## 10/Contro   10   15
## 2/Active    17   17
## 2/Contro     9   19
## 3/Active    15   24
## 3/Contro    20   13
## 4/Active    15   16
## 4/Contro    13   11
## 5/Active    21   25
## 5/Contro    16   19
## 6/Active    18   13
## 6/Contro    16   15
## 7/Active    21   16
## 7/Contro    17   14
## 8/Active    16   23
## 8/Contro    12    7
## 9/Active    17   24
## 9/Contro    11   25
```

The data set is quite unbalanced.

Fit a split plot model.

```
pa.lme <- lmer(logmvpa ~ block.f + school.f +
  gender.f + gender.f:school.f +
  (1| block.f:school.f), data=pa)
summary(pa.lme)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: logmvpa ~ block.f + school.f + gender.f + gender.f:school.f +
## (1 | block.f:school.f)
## Data: pa
##
## REML criterion at convergence: 1271.2
##
## Scaled residuals:
## Min 1Q Median 3Q Max
## -3.0776 -0.6658 -0.0041 0.6723 3.4756
##
## Random effects:
## Groups Name Variance Std.Dev.
## block.f:school.f (Intercept) 0.0175 0.1323
## Residual 0.3807 0.6170
## Number of obs: 658, groups: block.f:school.f, 20
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 1.918530 0.130127 14.744
## block.f2 -0.580533 0.171626 -3.383
## block.f3 -0.472930 0.169344 -2.793
## block.f4 -0.959434 0.174050 -5.512
## block.f5 0.429032 0.167477 2.562
## block.f6 0.363640 0.171544 2.120
## block.f7 -0.226452 0.170074 -1.331
## block.f8 -0.019471 0.174472 -0.112
## block.f9 -0.238797 0.168117 -1.420
## block.f10 -0.287648 0.173245 -1.660
## school.fContro -0.310697 0.092305 -3.366
## gender.fMale 0.150829 0.065303 2.310
## school.fContro:gender.fMale -0.004407 0.098256 -0.045
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it
```

Here's the type III ANOVA table.

```
pa.emm <- emmeans(pa.lme, c('school.f', 'gender.f'))
joint_tests(pa.emm)
```

```
## model term df1 df2 F.ratio p.value
## school.f 1 8.95 16.599 0.0028
## gender.f 1 643.54 9.136 0.0026
## school.f:gender.f 1 642.55 0.002 0.9643
```

How big is the difference between intervention and control?

```
pa.school <- emmeans(pa.lme, 'school.f')
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
pa.school
```

```
## school.f emmean SE df lower.CL upper.CL
## Active 1.79 0.0531 8.20 1.67 1.92
## Contro 1.48 0.0555 9.78 1.36 1.61
##
```

```
## Results are averaged over the levels of: block.f, gender.f
```

```
## Degrees-of-freedom method: kenward-roger
```

```
## Confidence level used: 0.95
```

```
summary(contrast(pa.school, list(diff=c(1, -1))),
infer=c(T,T))
```

```
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## diff 0.313 0.0768 8.95 0.139 0.487 4.074 0.0028
##
```

```
## Results are averaged over the levels of: block.f, gender.f
```

```
## Degrees-of-freedom method: kenward-roger
```

```
## Confidence level used: 0.95
```

And between genders?

```
pa.gender <- emmeans(pa.lme, 'gender.f')
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
pa.gender
```

```
## gender.f emmean SE df lower.CL upper.CL
## Female 1.56 0.0462 18.7 1.47 1.66
## Male 1.71 0.0450 16.7 1.62 1.81
##
```

```
## Results are averaged over the levels of: block.f, school.f
```

```
## Degrees-of-freedom method: kenward-roger
```

```
## Confidence level used: 0.95
```

```
summary(contrast(pa.gender, list(diff=c(1, -1))),
infer=c(T,T))
```

```
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## diff -0.149 0.0492 644 -0.245 -0.0521 -3.023 0.0026
##
```

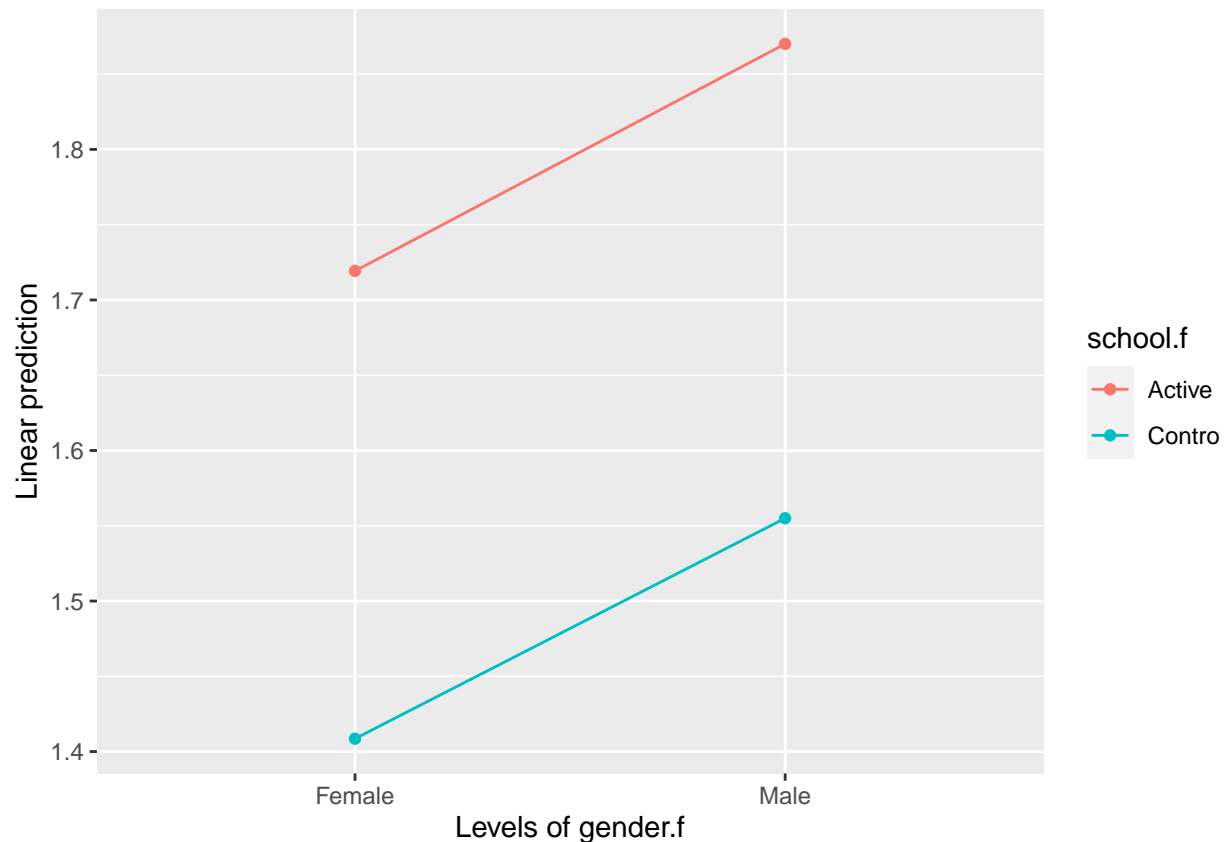
```
## Results are averaged over the levels of: block.f, school.f
```

```
## Degrees-of-freedom method: kenward-roger
```

```
## Confidence level used: 0.95
```

If the interaction were significant, I would explore comparisons between each treatment/gender group, probably starting with a plot.

```
emmip(pa.emm, school.f ~ gender.f)
```



What if the main plot design is a CRD? You still need to indicate the main plot error. This requires a variable that indicates each school. I'll demonstrate using the PA data, i.e. ignoring blocks. Bs is the variable that uniquely identifies each school. R is able to figure out that bs is nested in school.

```
pa2$bs.f <- factor(pa2$bs)
pa2.lme2 <- lmer(logmvpa ~ school.f +
  gender.f + gender.f:school.f +
  (1| bs.f), data=pa2)
summary(pa2.lme2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: logmvpa ~ school.f + gender.f + gender.f:school.f + (1 | bs.f)
## Data: pa2
##
## REML criterion at convergence: 418.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.52637 -0.67517  0.05858  0.71452  3.03319
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## bs.f     (Intercept)          0.206     0.4538
## Residual                            0.386     0.6213
```

```

## Number of obs: 200, groups:  bs.f, 20
##
## Fixed effects:
##
##              Estimate Std. Error t value
## (Intercept)      1.7602    0.1683  10.460
## school.fContro  -0.2666    0.2380  -1.120
## gender.fMale      0.1696    0.1243   1.365
## school.fContro:gender.fMale -0.1948    0.1757  -1.109
##
## Correlation of Fixed Effects:
##          (Intr) schl.C  gndr.M
## schol.fCntr -0.707
## gender.fMal -0.369  0.261
## schl.fCn:.M  0.261 -0.369 -0.707

```