

Meadowfoam Example: ANCOVA with 2 groups

This data is taken from Ramsey and Schafer. The response variable is number of flowers per plant. The predictors are light intensity (continuous) and time (Before or At PFI, categorical).

The model matrices are shown for illustration, but are not of interest for typical analysis. We use the emmeans package to get comparisons for time (cat predictor) adjusted for light.

We consider several models:

1. Model1: No interaction model forces the slope to be the same for both groups.
2. Model2: The interaction model allows different slopes for the two groups. This model is good for testing, but requires some work to get the estimated intercepts and slopes for the groups.
3. Model3: Alternate parameterization of the previous model (allowing different slopes for the two groups). This model allows the intercept and slope estimates to be obtained directly.

Just looking at the summary graph, the no interaction model seems reasonable. This is supported by the non-significant interaction in Model2.

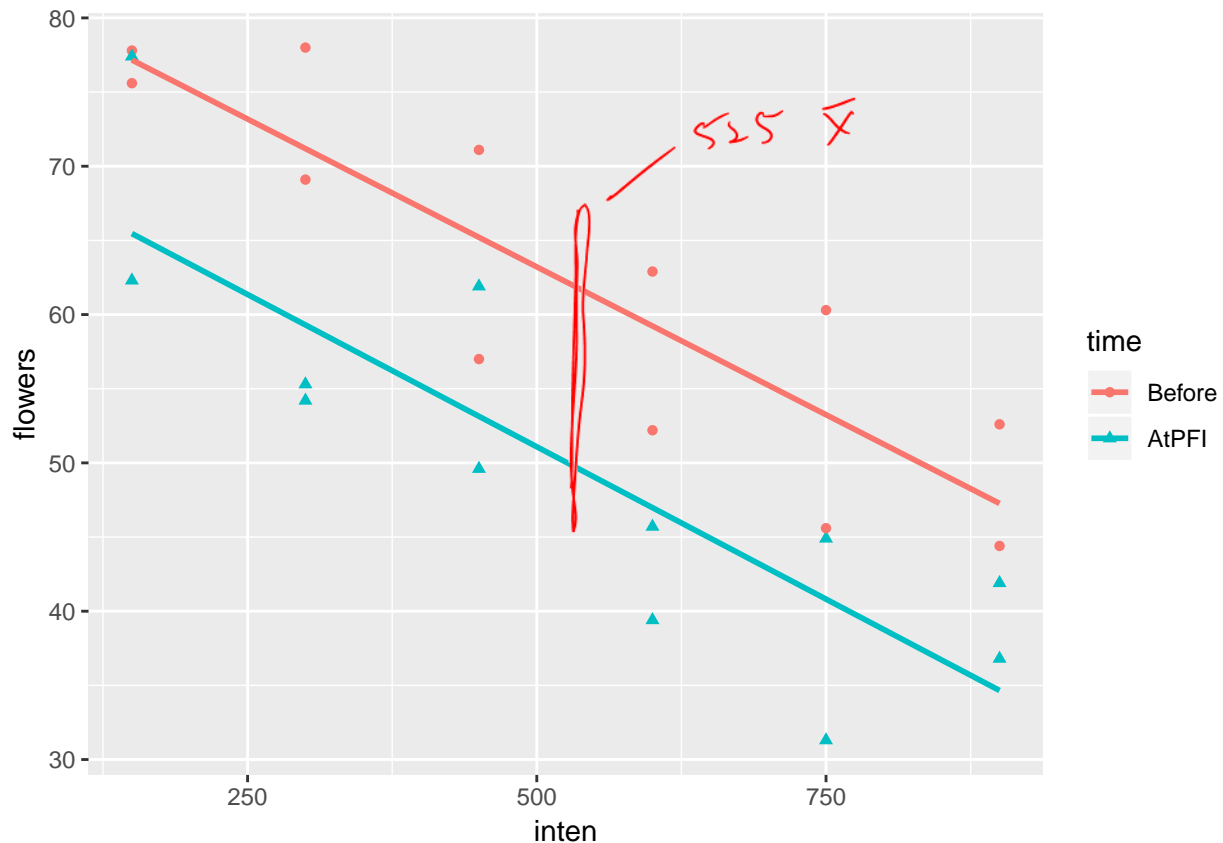
```
library(ggplot2)
library(car)
library(emmeans)
Meadow <- read.csv("~/Dropbox/STAT512/Lectures/MultReg2/MR2_Meadowfoam.csv")
str(Meadow)
```

```
## 'data.frame': 24 obs. of 4 variables:
## $ rep : int 1 1 1 1 1 1 1 1 1 ...
## $ flowers: num 62.3 55.3 49.6 39.4 31.3 36.8 77.8 69.1 57 62.9 ...
## $ time : Factor w/ 2 levels "AtPFI","Before": 1 1 1 1 1 1 2 2 2 ...
## $ inten : int 150 300 450 600 750 900 150 300 450 600 ...
```

```
#Reorder factor levels to have Before come first!
Meadow$time <- factor(Meadow$time, levels(Meadow$time)[c(2,1)])
levels(Meadow$time)
```

```
## [1] "Before" "AtPFI"
```

```
p <- qplot(inten, flowers, shape = time, color = time, data = Meadow)
p + geom_smooth(method = "lm", se = FALSE)
```



```
#Average flowers and inten by time
aggregate(cbind(flowers, inten) ~ time, data = Meadow, FUN = mean)

##      time  flowers  inten
## 1 Before 62.21667   525
## 2 AtPFI 50.05833   525
```

Model1: ANCOVA NO Interaction

Model.matrix not usually of direct interest, shown here for illustration. For this example (designed experiment), the emmeans are the same as the simple means. The emmeans represent the predicted response at the average value of the continuous predictor (inten = 525) in this case.

```
Model1 <- lm(flowers ~ time + inten, data = Meadow)
model.matrix(Model1)
```

```
##      (Intercept) timeAtPFI inten
## 1             1         1   150
## 2             1         1   300
## 3             1         1   450
## 4             1         1   600
## 5             1         1   750
## 6             1         1   900
## 7             1         0   150
## 8             1         0   300
## 9             1         0   450
## 10            1         0   600
```

Handwritten red notes include 'At' with an arrow pointing to the 'timeAtPFI' column and 'Before' with an arrow pointing to the 'inten' column.

```
## 11      1      0  750
## 12      1      0  900
## 13      1      1  150
## 14      1      1  300
## 15      1      1  450
## 16      1      1  600
## 17      1      1  750
## 18      1      1  900
## 19      1      0  150
## 20      1      0  300
## 21      1      0  450
## 22      1      0  600
## 23      1      0  750
## 24      1      0  900
```

```
## attr("assign")
## [1] 0 1 2
## attr("contrasts")
## attr("contrasts")$time
## [1] "contr.treatment"
```

```
summary(Model1)
```

```
##
## Call:
## lm(formula = flowers ~ time + inten, data = Meadow)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.652 -4.139 -1.558  5.632 12.165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  83.464167   3.273772  25.495 < 2e-16 ***
## timeAtPFI    -12.158333   2.629557  -4.624 0.000146 ***
## inten        -0.040471   0.005132  -7.886 1.04e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.441 on 21 degrees of freedom
## Multiple R-squared:  0.7992, Adjusted R-squared:  0.78
## F-statistic: 41.78 on 2 and 21 DF, p-value: 4.786e-08
```

```
Anova(Model1, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: flowers
##              Sum Sq Df F value    Pr(>F)
## (Intercept) 26966.2  1 649.985 < 2.2e-16 ***
## time         887.0  1  21.379 0.000146 ***
## inten       2579.8  1  62.181 1.037e-07 ***
## Residuals    871.2 21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
emmeans(Model1, pairwise ~ time)
```

```
## $emmeans
##   time      emmean      SE df lower.CL upper.CL
## Before 62.21667 1.859378 21 58.34988 66.08345
## AtPFI  50.05833 1.859378 21 46.19155 53.92512
##
## Confidence level used: 0.95
##
## $contrasts
##   contrast      estimate      SE df t.ratio p.value
## Before - AtPFI 12.15833 2.629557 21  4.624 0.0001
```

```
NewData <- data.frame(inten = rep(525, 2), time = c("Before", "AtPFI"))
predict(Model1, newdata = NewData)
```

```
##      1      2
## 62.21667 50.05833
```

same as .0001

Model2: ANCOVA WITH Interaction #1

Model.matrix not usually of direct interest, shown here for illustration. The emmeans() function returns a warning/note when the model includes interaction but we ask for a comparison of "main effects". This is not a problem, just need to understand what's being tested.

```
Model2 <- lm(flowers ~ time*inten, data = Meadow)
model.matrix(Model2)
```

```
##      (Intercept) timeAtPFI inten timeAtPFI:inten
## 1             1             1    150             150
## 2             1             1    300             300
## 3             1             1    450             450
## 4             1             1    600             600
## 5             1             1    750             750
## 6             1             1    900             900
## 7             1             0    150              0
## 8             1             0    300              0
## 9             1             0    450              0
## 10            1             0    600              0
## 11            1             0    750              0
## 12            1             0    900              0
## 13            1             1    150             150
## 14            1             1    300             300
## 15            1             1    450             450
## 16            1             1    600             600
## 17            1             1    750             750
## 18            1             1    900             900
## 19            1             0    150              0
## 20            1             0    300              0
## 21            1             0    450              0
## 22            1             0    600              0
## 23            1             0    750              0
## 24            1             0    900              0
## attr(,"assign")
```

not added
to data frame
but from lm()

off for
Before

```
## [1] 0 1 2 3
## attr("contrasts")
## attr("contrasts")$time
## [1] "contr.treatment"
```

```
summary(Model2)
```

```
##
## Call:
## lm(formula = flowers ~ time * inten, data = Meadow)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.516 -4.276 -1.422  5.473 11.938
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    83.14667    4.343305   19.144 2.49e-14 ***
## timeAtPFI     -11.52333    6.142360   -1.876  0.0753 .
## inten         -0.039867    0.007435   -5.362 3.01e-05 ***
## timeAtPFI:inten -0.001210    0.010515   -0.115  0.9096
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.598 on 20 degrees of freedom
## Multiple R-squared:  0.7993, Adjusted R-squared:  0.7692
## F-statistic: 26.55 on 3 and 20 DF,  p-value: 3.549e-07
```

```
Anova(Model2, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: flowers
##              Sum Sq Df F value    Pr(>F)
## (Intercept) 15953.9  1 366.4790 2.486e-14 ***
## time         153.2  1   3.5195  0.07532 .
## inten       1251.6  1  28.7509 3.008e-05 ***
## time:inten    0.6  1   0.0132  0.90957
## Residuals    870.7 20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
emmeans(Model2, pairwise ~ time)
```

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

```
## $emmeans
##      time      emmean      SE df lower.CL upper.CL
## Before 62.21667 1.904665 20 58.24360 66.18973
## AtPFI  50.05833 1.904665 20 46.08527 54.03140
##
## Confidence level used: 0.95
##
## $contrasts
##      contrast      estimate      SE df t.ratio p.value
## Before - AtPFI 12.15833 2.693603 20   4.514 0.0002
```

at $\bar{y} = 525$ inten

```
predict(Model2, newdata = NewData)
```

```
##           1           2
## 62.21667 50.05833
```

Model3: ANCOVA WITH Interaction #2 Alternate Parameterization

This model is helpful for getting the slope and intercept estimates directly. But do NOT use the ANOVA table for this alternate parameterization!

```
Model3 <- lm(flowers ~ time + time:inten - 1, data = Meadow)
model.matrix(Model3)
```

```
##      timeBefore timeAtPFI timeBefore:inten timeAtPFI:inten
## 1           0           1              0              150
## 2           0           1              0              300
## 3           0           1              0              450
## 4           0           1              0              600
## 5           0           1              0              750
## 6           0           1              0              900
## 7           1           0             150               0
## 8           1           0             300               0
## 9           1           0             450               0
## 10          1           0             600               0
## 11          1           0             750               0
## 12          1           0             900               0
## 13          0           1              0              150
## 14          0           1              0              300
## 15          0           1              0              450
## 16          0           1              0              600
## 17          0           1              0              750
## 18          0           1              0              900
## 19          1           0             150               0
## 20          1           0             300               0
## 21          1           0             450               0
## 22          1           0             600               0
## 23          1           0             750               0
## 24          1           0             900               0
## attr("assign")
## [1] 1 1 2 2
## attr("contrasts")
## attr("contrasts")$time
## [1] "contr.treatment"
```

```
summary(Model3)
```

```
##
## Call:
## lm(formula = flowers ~ time + time:inten - 1, data = Meadow)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -9.516 -4.276 -1.422 5.473 11.938
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## timeBefore      83.146667   4.343305  19.144 2.49e-14 ***
## timeAtPFI       71.623333   4.343305  16.491 4.14e-13 ***
## timeBefore:inten -0.039867   0.007435  -5.362 3.01e-05 ***
## timeAtPFI:inten  -0.041076   0.007435  -5.525 2.08e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.598 on 20 degrees of freedom
## Multiple R-squared:  0.9891, Adjusted R-squared:  0.9869
## F-statistic: 454.3 on 4 and 20 DF,  p-value: < 2.2e-16
```

```
Anova(Model3, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: flowers
##           Sum Sq Df F value    Pr(>F)
## time          27792.2  2 319.208 6.688e-16 ***
## time:inten    2580.3  2  29.636 1.045e-06 ***
## Residuals      870.7 20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

before intercept
at

slopes

don't use -1
this w/ intercept
parameterization