Homework #2

Dustin Leatherman

January 19, 2019

Meadowfoam and Time

$\mu\{\text{Flowers}|\text{light,time}\} = \gamma_0 + \gamma_1\text{light} + \gamma_2\text{time}$

a

$\mu\{\text{Flowers}|\text{light, time=0}\} = \gamma_0 + \gamma_1\text{light}$ $\mu\{\text{Flowers}|\text{light, time=24}\} = \gamma_0 + \gamma_1\text{light} + 24\gamma_2 \rightarrow (\gamma_0 + 24\gamma_2) + \gamma_1\text{light}$

When time is changed to a continuous variable, it affects the Intercept. The intercept will be increased by $24\gamma_2$ for flowers who started light treatment 24 days before prior to PFI.

b

$\mu\{\text{Flowers}|\text{light, early=0}\} = \beta_0 + \beta_1\text{light}$ $\mu\{\text{Flowers}|\text{light, early=1}\} = (\beta_0 + \beta_2) + \beta_1\text{light}$

When early=0 is substituted for time=0, the intercepts are $\beta_1 = \gamma_1$. When early=1 is substituted for time=24, $\beta_2 = 24\gamma_2$.

c

```r
case0901$early <- ifelse(case0901$Time==2,1,0)
case0901$time.cont <- ifelse(case0901$Time==2,24,0)

model.ind <- lm(Flowers ~ Intensity + early, data=case0901)
summary(model.ind)
```

```
##
## Call:
## lm(formula = Flowers ~ Intensity + early, data = case0901)
##
## 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) 71.305833   3.273772  21.781 6.77e-16 ***
## Intensity   -0.040471   0.005132  -7.886 1.04e-07 ***
## early       12.158333   2.629557   4.624 0.000146 ***
## ---
## 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
```

```r
model.time <- lm(Flowers ~ Intensity + time.cont, data=case0901)
summary(model.time)
```

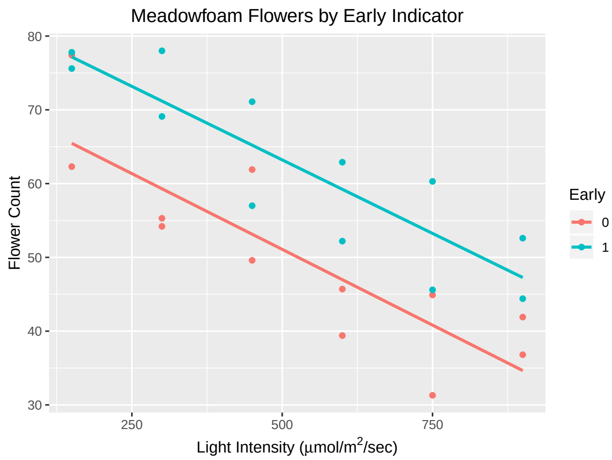
```
##
## Call:
## lm(formula = Flowers ~ Intensity + time.cont, data = case0901)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.652 -4.139 -1.558  5.632 12.165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 71.305833   3.273772  21.781 6.77e-16 ***
## Intensity   -0.040471   0.005132  -7.886 1.04e-07 ***
## time.cont    0.506597   0.109565   4.624 0.000146 ***
## ---
## 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
```

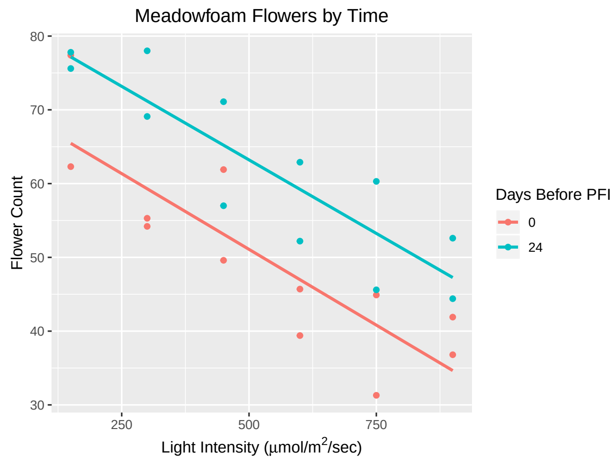
The estimated effect on mean Flower count for early=1 is 12.158 additional flowers for the indicator variable model while the estimated effect on mean Flower count for each day before PFI is 0.507. When time=24 and early=1, the same Flower count is achieved. For both models, the estimated mean flower count based on Intensity are equivalent which indicates that the slope is consistent between the two. By using a continuous variable instead of an indicator variable, the results end up being the same but it makes analysis easier to draw estimated Flower counts for values between 0 and 24 days prior to PFI. This can be verified by comparing the slopes and intercepts in the graphs of each model.

```r
early.plot <- ggplot(case0901, aes(color = factor(early), x = Intensity, y = Flowers)) +
  geom_point() +
  geom_smooth(method="lm", se=F) +
  xlab(expression(paste("Light Intensity (", mu, "mol/",m^2, "/sec)"))) +
  ylab("Flower Count") +
  labs(color=guide_legend("Early")) +
  ggtitle("Meadowfoam Flowers by Early Indicator") +
    theme(plot.title = element_text(hjust = 0.5))

time.plot <- ggplot(case0901, aes(color = factor(time.cont), x = Intensity, y = Flowers)) +
  geom_point() +
  geom_smooth(method="lm", se=F) +
  xlab(expression(paste("Light Intensity (", mu, "mol/",m^2, "/sec)"))) +
  ylab("Flower Count") +
  labs(color=guide_legend("Days Before PFI")) +
  ggtitle("Meadowfoam Flowers by Time") +
    theme(plot.title = element_text(hjust = 0.5))

grid.arrange(early.plot, time.plot,
               widths = c(1,1),
               ncol = 2)
```





Intensity By Indicator Variables

$\mu\{\text{Flowers}|\text{light,early}\} = \beta_0 + \beta_1\text{L300} + \beta_2\text{L450} + \beta_3\text{L600} + \beta_4\text{L750} + \beta_5\text{L900} + \beta_6\text{early}$

a

$\mu\{\text{Flowers}|\text{light,L300} = \text{L450} = \text{L600} = \text{L750} = \text{L900} = 0\text{, early} = 1\} = \beta_0 + \beta_6$

$\mu\{\text{Flowers}|\text{light,early} = \text{L300} = \text{L450} = \text{L600} = \text{L750} = \text{L900} = 0\} = \beta_0$

$\mu\{\text{Flowers}|\text{light,L300} = \text{L600} = \text{L750} = \text{L900} = 0\text{, early} = \text{L450} = 1\} = \beta_0 + \beta_2 + \beta_6$

$\mu\{\text{Flowers}|\text{light,early} = \text{L300} = \text{L600} = \text{L750} = \text{L900} = 0\text{, L450} = 1\} = \beta_0 + \beta_2$

b

$\mu\{\text{Flowers}|\text{light,early}\} = \beta_0 + \beta_1\text{L300} + \beta_2\text{L450} + \beta_3\text{L600} + \beta_4\text{L750} + \beta_5\text{L900} + \beta_6\text{early} + \beta_7(\text{L300} \times \text{early}) + \beta_8(\text{L450} \times \text{early}) + \beta_9(\text{L600} \times \text{early}) + \beta_1 0(\text{L750} \times \text{early}) + \beta_1 1(\text{L900} \times \text{early})$

$\mu\{\text{Flowers}|\text{light, L300} = \text{L450} = \text{L600} = \text{L750} = \text{L900} = 0\text{, early} = 1\} = \beta_0 + \beta_6$

$\mu\{\text{Flowers}|\text{light,early} = \text{L300} = \text{L450} = \text{L600} = \text{L750} = \text{L900} = 0\} = \beta_0$

$\mu\{\text{Flowers}|\text{light,L300} = \text{L600} = \text{L750} = \text{L900} = 0\text{, early} = \text{L450} = 1\} = \beta_0 + \beta_2 + \beta_6 + \beta_8$

$\mu\{\text{Flowers}|\text{light,early} = \text{L300} = \text{L600} = \text{L750} = \text{L900} = 0\text{, L450} = 1\} = \beta_0 + \beta_2$

c

For the model where there are no interaction terms, early affects the mean number of flowers by β_6. For the model with interaction terms, early affects the mean number of flowers by β_6 when light intensity is 150 μmol/m^2/sec. When early and the light intensity exceeds 150 μmol/m^2/sec, the mean number of flowers increases by $\beta_6 + \beta_x$ associated with the interaction term. For example, the mean number of flowers increases by $\beta_6 + \beta_8$ when a 450 μmol/m^2/sec light treatment is used.

Graph A is associated with the model with no interactions because the distance between early and late is consistent between each of the intensities. Graph B is associated with the model with interactions since the distance between early and late varies between intensities.

Depression

a

The following indicator variables represent the education categories described in the survey: SOMECOLLEGE, COLLEGE. The baseline indicator is the Highschool category.

$\mu\{\text{Score}|\text{age, SOMECOLLEGE, COLLEGE}\} = \beta_0 + \beta_1\text{age} + \beta_2\text{SOMECOLLEGE} + \beta_3\text{COLLEGE} + \beta_4(\text{SOMECOLLEGE} \times \text{age}) + \beta_5(\text{COLLEGE} \times \text{age})$

$\mu\{\text{Score}|\text{age, SOMECOLLEGE} = 0\text{, COLLEGE} = 1\} = \beta_0 + \beta_1\text{age} + \beta_3 + \beta_5\text{age}$

$\mu\{\text{Score}|\text{age, COLLEGE} = \text{SOMECOLLEGE} = 0\} = \beta_0 + \beta_1\text{age}$

$\rightarrow \beta_0 + \beta_1\text{age} + \beta_3 + \beta_5\text{age - } (\beta_0 + \beta_1\text{age})$

$\rightarrow \beta_3 + \beta_5\text{age}$

The interaction terms may cause unequal slopes and intercepts while having age as a parameter alone will ensure that the score changes linearly. In this case, β_5 measures the diverging gap between those with a Highschool degree and those with a college degree.

b

$\mu\{\text{Score}|\text{age, SOMECOLLEGE, COLLEGE}\} = \beta_0 + \beta_1\text{age} + \beta_2\text{SOMECOLLEGE} + \beta_2\text{COLLEGE} + \beta_3(\text{SOMECOLLEGE} \times \text{age}) + \beta_3(\text{COLLEGE} \times \text{age})$

$\mu\{\text{Score}|\text{age, SOMECOLLEGE} = 0\text{, COLLEGE} = 1\} = \beta_0 + \beta_1\text{age} + \beta_2 + \beta_3\text{age}$

$\mu\{\text{Score}|\text{age, COLLEGE} = \text{SOMECOLLEGE} = 0\} = \beta_0 + \beta_1\text{age}$

$\rightarrow \beta_0 + \beta_1\text{age} + \beta_2 + \beta_3\text{age - } (\beta_0 + \beta_1\text{age})$

$\rightarrow \beta_2 + \beta_3\text{age}$

By using the same coefficients for SOMECOLLEGE and COLLEGE, the model will have the same slopes and intercepts for both categories. β_3 measures the divergence between the two categories in this model.