F-test model selection

Maxwel C. Oliveira

February 18, 2017

Create a R-studio file

At the R-studio toolbar, click in file, New Project..., Existing Directory, and choose the folder that contain you data. Your data file have to be in that chosen folder. In addition, we recommend your data file in csv (comma delimited) file.

Load the data in R-studio

Assign the name of your data set (replace "DMT" to a name of your choice). If you use comma (,) for separating decimals places use read.csv2(). If you use a decimal point (.), use read.csv(). In parenthesis, write the name of your data set file.

```
DMT=read.csv("dmtotal.csv")
```

The command head prints the first 6 lines of the data set. It is useful for double checking your data set.

```
head(DMT)
```

```
##   block treat densitycrop densityweed species1g        yl weed
## 1     1     1           1           0     76.98 -3.523400    1
## 2     2     1           1           0     69.98  5.890264    1
## 3     3     1           1           0     71.99  3.187197    1
## 4     4     1           1           0     77.14 -3.738569    1
## 5     1     2           1           1     23.76 68.047337    1
## 6     2     2           1           1     28.53 61.632598    1
```

The command str prints how R-studio is reading the characters vector in a data set.

```
str(DMT)
```

```
## 'data.frame':    40 obs. of  7 variables:
##  $ block      : int  1 2 3 4 1 2 3 4 1 2 ...
##  $ treat      : int  1 1 1 1 2 2 2 2 3 3 ...
##  $ densitycrop: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ densityweed: int  0 0 0 0 1 1 1 1 2 2 ...
##  $ species1g  : num  77 70 72 77.1 23.8 ...
##  $ yl         : num  -3.52 5.89 3.19 -3.74 68.05 ...
##  $ weed       : int  1 1 1 1 1 1 1 1 1 1 ...
```

Fit a full model, Cousens model with 4 parameters.

Full is a user-defined name that will contain all information about the fitted model generated by nls function. Start is used to estimate values of parameter I and A in the model. It can easier be determined from visual inspection of the data set. The bracket [] after each parameter in the equation tell R to estimate a parameter for each weed species (4 parameters).

```
Full = nls(yl ~ (I[weed]*densityweed)/(1+(I[weed]/A[weed])*densityweed),
data=DMT, start=list(I=c(60,30), A=c(80,60)), trace=T)
```

```
## 31213.15 :  60 30 80 60
## 2425.918 :  145.75104  51.26517  99.25544 118.03099
## 1490.876 :  167.38375  51.78702 107.96366 118.03773
## 1490.847 :  167.21379  51.76884 108.06086 118.07677
## 1490.847 :  167.22908  51.77015 108.05823 118.07434
```

Check estimated parameters.

This command provides the estimated parameters I and A for each weed species.

```
summary(Full)
```

```
##
## Formula: yl ~ (I[weed] * densityweed)/(1 + (I[weed]/A[weed]) *
densityweed)
##
## Parameters:
##    Estimate Std. Error t value Pr(>|t|)
## I1  167.229     29.843   5.604 2.35e-06 ***
## I2   51.770      7.792   6.644 9.66e-08 ***
## A1  108.058      5.657  19.102  < 2e-16 ***
## A2  118.074     14.640   8.065 1.39e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.435 on 36 degrees of freedom
##
## Number of iterations to convergence: 4
## Achieved convergence tolerance: 5.431e-06
```

Fit a reduced model (Red.1), Cousens model with 2 parameters.

Red.1 is a user-defined name that will contain information about the first reduced model generated by nls function. Notice that we do not include bracket [] after each parameter I and A. In this case we are combining parameter I and A for both weed species. Our hypothesis is that a single parameter I and A for both species is enough to describe the crop-weed relationship.

```
Red.1 = nls(yl ~ (I*densityweed)/(1+(I/A)*densityweed), data=DMT,
start=list(I=40, A=80), trace=T)

## 32191.27 :  40 80
## 7326.399 :  89.40880 98.09019
## 6282.934 :   95.51888 107.39302
## 6282.929 :   95.6256 107.3586
## 6282.929 :   95.6207 107.3611
```

Check estimated parameters.

This command provides the estimated parameters I and A for both weed species together.

```
summary(Red.1)

##
## Formula: yl ~ (I * densityweed)/(1 + (I/A) * densityweed)
##
## Parameters:
##   Estimate Std. Error t value Pr(>|t|)
## I    95.62      21.51   4.446 7.37e-05 ***
## A   107.36      11.10   9.674 8.51e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.86 on 38 degrees of freedom
##
## Number of iterations to convergence: 4
## Achieved convergence tolerance: 1.965e-06
```

Test a first hypothesis

Testing hypothesis using anova. We are using the Full model (separated I and A for each species) to compare with Red.1 (single I and A for both species). If p>0.05, we should use the Red.1 model, which means that the simplest model (Red.1) is enough to describe crop-weed relationship. If not we should proceed to the next hypothesis testing.

```
anova(Full, Red.1)

## Analysis of Variance Table
##
## Model 1: yl ~ (I[weed] * densityweed)/(1 + (I[weed]/A[weed]) *
densityweed)
## Model 2: yl ~ (I * densityweed)/(1 + (I/A) * densityweed)
##   Res.Df Res.Sum Sq Df  Sum Sq F value    Pr(>F)
## 1     36     1490.8
## 2     38     6282.9 -2 -4792.1  57.858 5.687e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-test showed P<0.05, therefore the Red.1 model is not enought to describe the crop-weed relationship.

Fit a reduced model (Red.2), Cousens model with 3 parameters.

Red.2 is a user-defined name that will contain information about the second reduced model generated by nls function. Notice that the bracket [] is after the parameter A only, which means that we are testing a hypothesis of single parameter I, but different A for the species.

```
Red.2 = nls(yl ~ (I*densityweed)/(1+(I/A[weed])*densityweed), data=DMT,
start=list(I=60, A=c(80,60)),  trace=T)
```

```
## 24826.57 :  60 80 60
## 2585.506 :  105.76104 124.72924  81.52309
## 2569.254 :  112.12412 122.30136  80.95197
## 2568.845 :  112.01615 122.68514  80.92959
## 2568.842 :  112.14466 122.62227  80.90443
## 2568.842 :  112.13957 122.62725  80.90283
## 2568.842 :  112.14182 122.62612  80.90242
```

Check estimated parameters.

This command provides the estimated parameters I for both weed species and A for each weed species.

```
summary(Red.2)
```

```
##
## Formula: yl ~ (I * densityweed)/(1 + (I/A[weed]) * densityweed)
##
## Parameters:
##    Estimate Std. Error t value Pr(>|t|)
## I   112.142     17.347   6.465 1.48e-07 ***
## A1  122.626      8.833  13.882 3.03e-16 ***
## A2   80.902      5.058  15.995  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.332 on 37 degrees of freedom
##
## Number of iterations to convergence: 6
## Achieved convergence tolerance: 2.932e-06
```

Test a second hypothesis

Testing hypothesis using F-test. We are using the Full model (separated I and A for each species) to compare with Red.2 (single I and different A for both species). If p>0.05, we should use the Red.2 model, which means that the simplest model (Red2) is enough to describe crop-weed relationship. If not we should proceed to the next hypothesis testing.

```
## Analysis of Variance Table
##
## Model 1: yl ~ (I[weed] * densityweed)/(1 + (I[weed]/A[weed]) *
densityweed)
## Model 2: yl ~ (I * densityweed)/(1 + (I/A[weed]) * densityweed)
##   Res.Df Res.Sum Sq Df Sum Sq F value  Pr(>F)
## 1     36     1490.8
## 2     37     2568.8 -1  -1078  26.031 1.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-test showed P<0.05, therefore the Red.2 model is not enought to describe the crop-weed relationship.

Fit a reduced model (Red.3), 3 parameters.

Red.3 is a user-defined name that will contain information about the third reduced model generated by nls function. Notice that the bracket [] is after the parameter I only, which means that we are testing a hypothesis of different parameter I, but single parameter A for the species.

```
Red.3 = nls(yl ~ (I[weed]*densityweed)/(1+(I[weed]/A)*densityweed), data=DMT,
start=list(I=c(30,30), A=70), trace=T)

## 48504.32 :  30 30 70
## 6826.593 :  108.26052  64.67140  90.67584
## 1591.007 :  156.88369  51.99749 109.51464
## 1511.456 :  160.65270  56.41664 109.53960
## 1511.333 :  159.86393  56.42673 109.73773
## 1511.333 :  159.8977  56.4320 109.7317
## 1511.333 :  159.89585  56.43162 109.73207
```

Check estimated parameters.

This command provides the estimated parameters I for each weed species and A for both weed species.

```
summary(Red.3)

##
## Formula: yl ~ (I[weed] * densityweed)/(1 + (I[weed]/A) * densityweed)
##
## Parameters:
##    Estimate Std. Error t value Pr(>|t|)
## I1  159.896     25.373   6.302 2.45e-07 ***
## I2   56.432      4.915  11.482 9.31e-14 ***
## A   109.732      5.243  20.929  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.391 on 37 degrees of freedom
```

```
##
## Number of iterations to convergence: 6
## Achieved convergence tolerance: 9.501e-07
```

Test a third hypothesis

Testing hypothesis using F-test. We are using the Full model (separated I and A for each species) to compare with Red.3 (different I and single A for both species). If p>0.05, we should use the Red.3 model, which means that the simplest model (Red.3) is enough to describe crop-weed relationship.

```
anova(Full, Red.3)
```

```
## Analysis of Variance Table
##
## Model 1: yl ~ (I[weed] * densityweed)/(1 + (I[weed]/A[weed]) *
densityweed)
## Model 2: yl ~ (I[weed] * densityweed)/(1 + (I[weed]/A) * densityweed)
##   Res.Df Res.Sum Sq Df  Sum Sq F value Pr(>F)
## 1     36     1490.8
## 2     37     1511.3 -1 -20.487  0.4947 0.4864
```

Results showed that P >0.05. Therefore, the Full model can be simplified to Red.3 model.

Plotting the Red.3 model

The par command is used to define the plot size. The command plot and lines are used to generate the figure and the averaged points of yield loss at each density (Figure 4). The subset command is adding each weed species separately in the plot (weed 1) and lines (weed 2).

The x is a user-defined name, it will contain the x-axis sequence of the data set. weed1 and weed2 is also a user-defined name, this is the equation with the previous parameter estimates I and A estimated from Red.3 model using the nls function. Notice that the parameters estimated in Red.3 model were inserted in the Cousens model for each weed species.

The command lines will insert the previous equation into the plot. Command lty, lwd, and col define the line type, size, and color, respectively.

The command legend will add the legend into the previous created plot.

```
par(mar=c(5,6,2,2), mgp=c(4,1.5,0))
plot(yl~densityweed, data=DMT, subset = weed =="1", pch=16, font.size=2,
cex=1, las=1, xlab=expression("Weed Density (plants pot"^-1*")"), ylim=c(-
10,110), ylab = "Yield Loss (%)", cex.axis=1, cex.lab=1)
lines(yl~densityweed, type="p",data=DMT, subset = weed =="2", col=2, cex=1,
pch=1)

x=seq(0,4,0.25)
```

```
weed1=(159.896*x)/(1+(159.896/109.732)*x)
weed2=(56.432*x)/(1+(56.432/109.732)*x)

lines(x,weed1, lty=1, lwd=1, col=1)
lines(x,weed2, lty=3, lwd=1, col=2)

legend("bottomright", legend=c("C. benghalensis", "R. brasiliensis"),
text.font = 3, col=c(1,2), pch= c(16,1), lty=c(1,3), lwd= c(1,1), bty="n",
cex=1)
```