

glms for proportion abscised

E.E. Jackson

30 April, 2020

```
rm(list = ls())
require(tidyverse)
require(lme4)
library(knitr)
library(report)
opts_chunk$set(tidy.opts=list(width.cutoff=50),tidy=TRUE)
```

Read dataset

```
fruitDat <- read.csv("../output/tables/cleanData.csv")
```

generalized linear models

```
## maximal model ##

fullmod <- glm(cbind(abscised_seeds, viable_seeds) ~
  seedpredationrate + log10(seed_dry) + height_avg +
  cvseed + log10(cofruit) + bcireproductive +
  endocarp_investment + seedpred_pres + lifeform,
  data = na.omit(fruitDat), family = quasibinomial(link = logit))

summary(fullmod)

##
## Call:
## glm(formula = cbind(abscised_seeds, viable_seeds) ~ seedpredationrate +
##   log10(seed_dry) + height_avg + cvseed + log10(cofruit) +
##   bcireproductive + endocarp_investment + seedpred_pres + lifeform,
##   family = quasibinomial(link = logit), data = na.omit(fruitDat))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -171.604  -26.817  -14.411    9.802   175.243
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.5457204   7.2247912   0.906 0.367795
## seedpredationrate 1.7548128   1.7108682   1.026 0.308293
## log10(seed_dry)  0.4520733   0.1182701   3.822 0.000269 ***
## height_avg      0.0310380   0.0260509   1.191 0.237188
## cvseed          0.5949215   0.3104011   1.917 0.059046 .
##
```

```

## log10(cofruit)      -3.4449990  3.1095531  -1.108  0.271409
## bcireproductive     0.0014034  0.0007682   1.827  0.071649 .
## endocarp_investment 3.2019430  0.4555351   7.029  7.74e-10 ***
## seedpred_pres      -0.3021825  0.4142502  -0.729  0.467957
## lifeformShrub       -0.3995954  0.6864183  -0.582  0.562192
## lifeformTree        0.5047212  0.3938680   1.281  0.203933
## lifeformUnderstory  0.3269848  0.4203957   0.778  0.439100
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 3209.143)
##
## Null deviance: 2023902 on 87 degrees of freedom
## Residual deviance: 198572 on 76 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
anova(fullmod, test = "F")

## Analysis of Deviance Table
##
## Model: quasibinomial, link: logit
##
## Response: cbind(abscised_seeds, viable_seeds)
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev      F    Pr(>F)
## NULL                                87    2023902
## seedpredationrate    1    599710      86    1424192 186.8753 < 2.2e-16 ***
## log10(seed_dry)      1    561068      85     863124 174.8342 < 2.2e-16 ***
## height_avg           1     73258      84     789866  22.8280 8.450e-06 ***
## cvseed               1    120986      83     668880  37.7003 3.477e-08 ***
## log10(cofruit)       1    121554      82     547326  37.8773 3.273e-08 ***
## bcireproductive      1     9271      81     538056   2.8889  0.09328 .
## endocarp_investment  1    332622      80     205434 103.6482 7.558e-16 ***
## seedpred_pres        1      256      79     205178   0.0798  0.77837
## lifeform             3     6606      76     198572   0.6862  0.56324
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
mod2 <- update(fullmod, ~. - lifeform) # drop variable with lowest p value

anova(fullmod, mod2, test = "F") # is the new model better?

## Analysis of Deviance Table
##
## Model 1: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
## height_avg + cvseed + log10(cofruit) + bcireproductive +
## endocarp_investment + seedpred_pres + lifeform
## Model 2: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
## height_avg + cvseed + log10(cofruit) + bcireproductive +
## endocarp_investment + seedpred_pres

```

```
##   Resid. Df Resid. Dev Df Deviance      F Pr(>F)
## 1       76      198572
## 2       79      205178 -3   -6606.1 0.6862 0.5632
```

```
# The simpler model is not significantly worse
```

```
anova(mod2, test = "F") # can we simplify further?
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model: quasibinomial, link: logit
```

```
##
```

```
## Response: cbind(abscised_seeds, viable_seeds)
```

```
##
```

```
## Terms added sequentially (first to last)
```

```
##
```

```
##
```

		Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
## NULL				87	2023902		
## seedpredationrate	1	599710		86	1424192	190.0811	< 2.2e-16 ***
## log10(seed_dry)	1	561068		85	863124	177.8335	< 2.2e-16 ***
## height_avg	1	73258		84	789866	23.2196	6.864e-06 ***
## cvseed	1	120986		83	668880	38.3470	2.491e-08 ***
## log10(cofruit)	1	121554		82	547326	38.5271	2.341e-08 ***
## bcireproductive	1	9271		81	538056	2.9384	0.09042 .
## endocarp_investment	1	332622		80	205434	105.4262	3.354e-16 ***
## seedpred_pres	1	256		79	205178	0.0811	0.77650

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod3 <- update(mod2, ~. - seedpred_pres)
```

```
anova(fullmod, mod2, mod3, test = "F")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
##           height_avg + cvseed + log10(cofruit) + bcireproductive +
```

```
##           endocarp_investment + seedpred_pres + lifeform
```

```
## Model 2: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
```

```
##           height_avg + cvseed + log10(cofruit) + bcireproductive +
```

```
##           endocarp_investment + seedpred_pres
```

```
## Model 3: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
```

```
##           height_avg + cvseed + log10(cofruit) + bcireproductive +
```

```
##           endocarp_investment
```

```
##   Resid. Df Resid. Dev Df Deviance      F Pr(>F)
```

```
## 1       76      198572
```

```
## 2       79      205178 -3   -6606.1 0.6862 0.5632
```

```
## 3       80      205434 -1    -256.0 0.0798 0.7784
```

```
anova(mod3, test = "F")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model: quasibinomial, link: logit
```

```
##
```

```
## Response: cbind(abscised_seeds, viable_seeds)
```

```
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev      F      Pr(>F)
## NULL                                87      2023902
## seedpredationrate      1      599710      86      1424192 194.7595 < 2.2e-16 ***
## log10(seed_dry)        1      561068      85      863124 182.2104 < 2.2e-16 ***
## height_avg             1      73258      84      789866  23.7911 5.368e-06 ***
## cvseed                 1     120986      83      668880  39.2908 1.735e-08 ***
## log10(cofruit)         1     121554      82      547326  39.4754 1.629e-08 ***
## bcireproductive        1       9271      81      538056   3.0107  0.08657 .
## endocarp_investment    1     332622      80      205434 108.0210 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

mod4 <- update(mod3, ~. - bcireproductive)
anova(fullmod, mod2, mod3, mod4, test = "F")

## Analysis of Deviance Table
##
## Model 1: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
##           height_avg + cvseed + log10(cofruit) + bcireproductive +
##           endocarp_investment + seedpred_pres + lifeform
## Model 2: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
##           height_avg + cvseed + log10(cofruit) + bcireproductive +
##           endocarp_investment + seedpred_pres
## Model 3: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
##           height_avg + cvseed + log10(cofruit) + bcireproductive +
##           endocarp_investment
## Model 4: cbind(abscised_seeds, viable_seeds) ~ seedpredationrate + log10(seed_dry) +
##           height_avg + cvseed + log10(cofruit) + endocarp_investment
##   Resid. Df Resid. Dev Df Deviance      F Pr(>F)
## 1         76      198572
## 2         79      205178 -3  -6606.1 0.6862 0.5632
## 3         80      205434 -1   -256.0 0.0798 0.7784
## 4         81      213771 -1  -8337.4 2.5980 0.1111

anova(mod4, test = "F")

## Analysis of Deviance Table
##
## Model: quasibinomial, link: logit
##
## Response: cbind(abscised_seeds, viable_seeds)
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev      F      Pr(>F)
## NULL                                87      2023902
## seedpredationrate      1      599710      86      1424192 197.800 < 2.2e-16 ***
## log10(seed_dry)        1      561068      85      863124 185.055 < 2.2e-16 ***
## height_avg             1      73258      84      789866  24.163 4.549e-06 ***
## cvseed                 1     120986      83      668880  39.904 1.354e-08 ***
```

```
## log10(cofruit)          1   121554          82      547326  40.092 1.270e-08 ***
## endocarp_investment    1   333555          81      213771 110.016 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mod4) # looks like this is our best
```

```
##
## Call:
## glm(formula = cbind(abscised_seeds, viable_seeds) ~ seedpredationrate +
##   log10(seed_dry) + height_avg + cvseed + log10(cofruit) +
##   endocarp_investment, family = quasibinomial(link = logit),
##   data = na.omit(fruitDat))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -181.827  -28.326  -13.629   9.535  150.425
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.51460     4.38161  -0.117  0.90680
## seedpredationrate  1.42439     1.26421   1.127  0.26320
## log10(seed_dry)    0.54746     0.07508   7.291 1.84e-10 ***
## height_avg        0.02686     0.01486   1.807  0.07450 .
## cvseed            0.73921     0.26371   2.803  0.00633 **
## log10(cofruit)    -0.45933     1.79745  -0.256  0.79895
## endocarp_investment 3.06175     0.28637  10.692 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 3031.894)
##
##      Null deviance: 2023902  on 87  degrees of freedom
## Residual deviance:  213771  on 81  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
report(mod4) # this gives us a nice interpretation of the model output
```

```
## Warning in eval(predvars, data, env): NaNs produced

## Warning in eval(predvars, data, env): NaNs produced
## Waiting for profiling to be done...

## Can't calculate log-loss.
## Can't calculate proper scoring rules for models without integer response values.
## `performance_pcp()` only works for models with binary response values.

## Warning in eval(predvars, data, env): NaNs produced

## Warning in eval(predvars, data, env): NaNs produced
## Waiting for profiling to be done...

## Can't calculate log-loss.
## Can't calculate proper scoring rules for models without integer response values.
```

```
## `performance_pcp()` only works for models with binary response values.  
## Can't calculate log-loss.  
## Can't calculate proper scoring rules for models without integer response values.  
## `performance_pcp()` only works for models with binary response values.  
  
## Warning in eval(predvars, data, env): NaNs produced  
  
## Warning in eval(predvars, data, env): NaNs produced  
  
## Waiting for profiling to be done...  
  
## We fitted a logistic model (estimated using ML) to predict cbind(abscised_seeds, viable_seeds) with :  
##  
## - The effect of seedpredationrate is positive and can be considered as small and not significant (beta = 0.7)  
## - The effect of log10(seed_dry) is positive and can be considered as small and significant (beta = 0.7)  
## - The effect of height_avg is positive and can be considered as very small and not significant (beta = 0.7)  
## - The effect of cvseed is positive and can be considered as very small and significant (beta = 0.7)  
## - The effect of log10(cofruit) is negative and can be considered as medium and not significant (beta = 0.7)  
## - The effect of endocarp_investment is positive and can be considered as very small and significant (beta = 0.7)  
##  
## - The effect of seedpredationrate is positive and can be considered as small and not significant (beta = 0.7)  
## - The effect of log10(seed_dry) is positive and can be considered as small and significant (beta = 0.7)  
## - The effect of height_avg is positive and can be considered as very small and not significant (beta = 0.7)  
## - The effect of cvseed is positive and can be considered as very small and significant (beta = 0.7)  
## - The effect of log10(cofruit) is negative and can be considered as medium and not significant (beta = 0.7)  
## - The effect of endocarp_investment is positive and can be considered as very small and significant (beta = 0.7)
```

Generalized linear mixed models

```
# read in the sp by year data
fruitDatFull <- read.csv("../output/tables/fullCleanData.csv")

# glmer won't take cbind, so we use weights= to
# weight by total_seeds.

# to account for temporal pseudoreplication the
# random-effects formula needs to indicate that the
# year (a continuous random effect) represents
# pseudoreplication within each individual sp? so
# (year/sp)

m1 <- lme4::glmer(proportion_abscised ~ seedpredationrate +
  (year | sp), family = binomial, data = fruitDatFull,
  weights = total_seeds)

## Warning in eval(family$initialize, rho): non-integer #successes in a binomial
## glm!

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.229236 (tol = 0.002, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable:
## - Rescale variables?;Model is nearly unidentifiable: large eigenvalue ratio
## - Rescale variables?

summary(m1)
```

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: proportion_abcised ~ seedpredationrate + (year | sp)
## Data: fruitDatFull
## Weights: total_seeds
##
##      AIC      BIC    logLik deviance df.resid
## 1120422.3 1120453.5 -560206.1 1120412.3     3777
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -214.651   -2.633   -0.441    2.137   289.300
##
## Random effects:
## Groups Name      Variance Std.Dev. Corr
## sp      (Intercept) 2.455e+04 156.67999
## year      6.128e-03  0.07828 -1.00
## Number of obs: 3782, groups: sp, 163
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.2013    0.1706  -7.042 1.90e-12 ***
## seedpredationrate  5.1169    1.1187   4.574 4.79e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr)
## seedprdtnt -0.417
## convergence code: 0
## Model failed to converge with max|grad| = 0.229236 (tol = 0.002, component 1)
## Model is nearly unidentifiable: very large eigenvalue
## - Rescale variables?
## Model is nearly unidentifiable: large eigenvalue ratio
## - Rescale variables?
# model won't converge

# try model based on data with means across years 1
# explanatory variable and 1 random effect

m2 <- lme4::glmer(proportion_abcised ~ seedpredationrate +
  (1 | sp), family = binomial, data = fruitDat, weights = total_seeds)

## Warning in eval(family$initialize, rho): non-integer #successes in a binomial
## glm!

summary(m2) #effect of seedpredationrate is significatly positive but very small

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: proportion_abcised ~ seedpredationrate + (1 | sp)
## Data: fruitDat

```

```

## Weights: total_seeds
##
##      AIC      BIC   logLik deviance df.resid
##  2160.3   2169.0 -1077.1   2154.3     134
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.70915 -0.01518  0.00214  0.01342  0.07793
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   sp      (Intercept) 2.404    1.55
## Number of obs: 137, groups:  sp, 137
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.2177    0.1693  -7.192  6.4e-13 ***
## seedpredationrate  4.6595    1.5534   3.000  0.0027 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr)
## seedprdtprt -0.615

```

```

m3 <- lme4::glmer(proportion_abscised ~ seedpred_pres +
  (1 | sp), family = binomial, data = fruitDat, weights = total_seeds)

```

```

## Warning in eval(family$initialize, rho): non-integer #successes in a binomial
## glm!

```

```

summary(m3)  #effect of seedpred_pres is signifcantly positive but very small

```

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial ( logit )
##   Formula: proportion_abscised ~ seedpred_pres + (1 | sp)
##   Data: fruitDat
##   Weights: total_seeds
##
##      AIC      BIC   logLik deviance df.resid
##  2161.2   2170.0 -1077.6   2155.2     134
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.75766 -0.01402  0.00235  0.01104  0.09809
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   sp      (Intercept) 2.433    1.56
## Number of obs: 137, groups:  sp, 137
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.5035    0.2508  -5.996 2.02e-09 ***

```



```

## seedpred_pres    0.8378      0.2968    2.823  0.00476 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## seedprd_prs -0.844

```