Ch1\_WTD\_modelDistribution

Erin Tattersall

May 8, 2018

Modelling detections rates between November 2015 and April 2018 for white-tailed deer.  
Deciding on most appropriate distribution for response variable.

Proportion of zeros in data

sum(det$WTDeer==0, na.rm = TRUE)/nrow(det)

## [1] 0.8877778

89% of the data is zeroes –> likely zero-inflated.

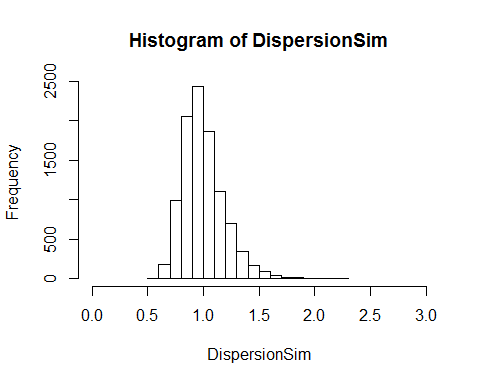
### Fitting a basic poisson GLM and checking overdispersion, using global model (doesn’t yet include SnowDays)

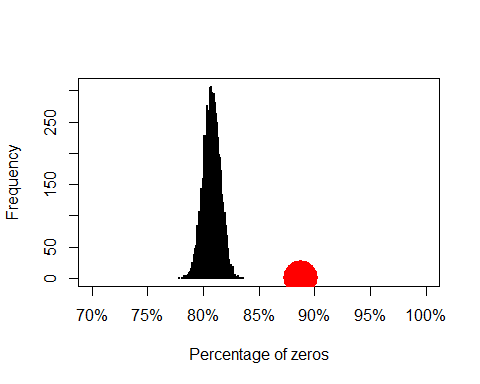
glm1 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays, data = det, family = poisson)  
# Residuals and overdispersion  
E1 <- resid(glm1, type="pearson")  
N <- nrow(det)  
p <- length(coef(glm1))  
sum(E1^2)/(N-p)

## [1] 2.380992

Dispersion value of 2.38 indicates overdispersion

Simulating data to check probability of calculated dispersion, given a Poisson GLM.

 Histogram indicates that an overdispersion statistic of 2.38 is well outside the likely distribution of dispersion statistics for Poisson distributed response variables, suggesting that WTDeer data is likely overdispersed.

Comparing the proportion of zeros in data to simulated zeros from model shows that WTDeer data has more zeroes than would be expected in a Poisson GLM  This presents a case for using zero-inflated models, which can be verified with model selection of GLMMs ## Model selection: choosing model form and distribution

Comparing the same GLMM (including random effects of Site and Month) modeled as a poisson, nb, ZIP and ZINB (with nbinom1 and nbinom2 differing in how variance changes with mean) yields:

## Model comparisons of distributions and zero-inflation  
glm1 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays + (1| Site) + (1|Month), data = det, family = poisson)  
glm2 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays + (1| Site) + (1|Month), data = det, family = nbinom1(link= "log"))  
glm3 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays + (1| Site) + (1|Month), data = det, family = nbinom2(link = "log"))  
glm4 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays + (1| Site) + (1|Month), zi = ~1, data = det, family = poisson)  
glm5 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays +(1| Site) + (1|Month), zi = ~1, data = det, family = nbinom1(link= "log")) #Warning 'matrix not positive definite'  
glm6 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays + (1| Site) + (1|Month), zi = ~1, data = det, family = nbinom2(link= "log")) #Warning 'matrix not positive definite'

## dLogLik dAIC df weight  
## Nbinom1 85.2 0.0 12 0.7087  
## ZINB1 85.3 1.8 13 0.2897  
## Nbinom2 78.8 12.8 12 0.0012  
## ZINB2 78.8 14.8 13 <0.001  
## ZIP 40.4 89.6 12 <0.001  
## Poisson 0.0 168.4 11 <0.001

ZINB1 models showed warnings (see above), but output includes parameter estimates, so models are included. Nbinom1 model output

## Family: nbinom1 ( log )  
## Formula:   
## WTDeer ~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays +   
## (1 | Site) + (1 | Month)  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 1435.7 1501.4 -705.8 1411.7 1758   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 0.9623 0.9810   
## Month (Intercept) 0.6894 0.8303   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom1 family (): 1.54   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.32095 1.22730 -1.891 0.05861 .   
## TreatmentHumanUse -0.30828 0.54948 -0.561 0.57478   
## TreatmentNatRegen -0.30620 0.47523 -0.644 0.51937   
## TreatmentSPP -1.36967 0.47354 -2.892 0.00382 \*\*   
## LineWidth -0.06730 0.12998 -0.518 0.60461   
## LD750 0.49909 0.22355 2.233 0.02558 \*   
## VegHt 0.28109 0.19532 1.439 0.15011   
## low2000 -5.81749 1.17719 -4.942 7.74e-07 \*\*\*  
## ActiveDays 0.12555 0.01536 8.176 2.95e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ZINB model output reveals that there is a significant probability of observing structured zeroes in WTDeer data

## Family: nbinom1 ( log )  
## Formula:   
## WTDeer ~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays +   
## (1 | Site) + (1 | Month)  
## Zero inflation: ~1  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 1437.5 1508.7 -705.7 1411.5 1757   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 0.9984 0.9992   
## Month (Intercept) 0.7178 0.8472   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom1 family (): 1.46   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.29986 1.24393 -1.849 0.06448 .   
## TreatmentHumanUse -0.31439 0.55766 -0.564 0.57291   
## TreatmentNatRegen -0.31815 0.48287 -0.659 0.50998   
## TreatmentSPP -1.38634 0.48072 -2.884 0.00393 \*\*   
## LineWidth -0.06759 0.13190 -0.512 0.60832   
## LD750 0.49787 0.22699 2.193 0.02828 \*   
## VegHt 0.27998 0.19856 1.410 0.15852   
## low2000 -5.88762 1.19726 -4.918 8.76e-07 \*\*\*  
## ActiveDays 0.12666 0.01551 8.165 3.20e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Zero-inflation model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.721 1.509 -2.466 0.0136 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Active Days affects the probability of observing a zero in data –> should be included in ZI model. It could be argued that it should NOT be included in conditional. I will test both

glm7 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 + ActiveDays +(1| Site) + (1|Month), zi = ~ActiveDays, data = det, family = nbinom1(link= "log"))  
glm8 <- glmmTMB(WTDeer~ Treatment + LineWidth + LD750 + VegHt + low2000 +(1| Site) + (1|Month), zi = ~ActiveDays, data = det, family = nbinom1(link= "log"))

## dLogLik dAIC df weight  
## ZINB1-AD1 94.9 0.0 13 1   
## Nbinom1 85.2 17.4 12 <0.001  
## ZINB1 85.3 19.2 13 <0.001  
## Nbinom2 78.8 30.2 12 <0.001  
## ZINB2 78.8 32.2 13 <0.001  
## ZIP 40.4 107.0 12 <0.001  
## Poisson 0.0 185.8 11 <0.001

## Family: nbinom1 ( log )  
## Formula:   
## WTDeer ~ Treatment + LineWidth + LD750 + VegHt + low2000 + (1 |   
## Site) + (1 | Month)  
## Zero inflation: ~ActiveDays  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 1418.3 1489.5 -696.1 1392.3 1757   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 1.0108 1.0054   
## Month (Intercept) 0.7048 0.8395   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom1 family (): 1.39   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.56122 1.15744 1.349 0.17738   
## TreatmentHumanUse -0.35737 0.55948 -0.639 0.52299   
## TreatmentNatRegen -0.36474 0.48421 -0.753 0.45129   
## TreatmentSPP -1.40393 0.48213 -2.912 0.00359 \*\*   
## LineWidth -0.06933 0.13234 -0.524 0.60035   
## LD750 0.50317 0.22792 2.208 0.02727 \*   
## VegHt 0.26205 0.19964 1.313 0.18930   
## low2000 -5.93209 1.19964 -4.945 7.62e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Zero-inflation model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 5.78699 1.46025 3.963 7.4e-05 \*\*\*  
## ActiveDays -0.36269 0.09845 -3.684 0.000229 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Including ActiveDays in the ZI model results in more logLikelihood of model and more AIC weight, with no warnings produced. For WTDeer models, I will therefore use zero inflated GLMMs with a nbinom1 distribution (where variance changes linearly with the mean), including ActiveDays in the ZI model.  
# Model hypotheses  
## Finding random structure

## dLogLik dAIC df weight  
## r2 70.3 0.0 16 1   
## rMonth 43.8 51.0 15 <0.001  
## rSite 19.5 99.5 15 <0.001  
## r0 0.0 136.5 14 <0.001

Continue modelling with 2 random effects. This model produced warnings, despite running in earlier comparison. Parameter estimates and SE obtained.