

BIOS6643. L18 Models for data with many zeros

Salamander data

A data set containing counts of salamanders with site covariates and sampling covariates. Each of 23 sites was sampled 4 times. See Price et al. (2016, Effects of mountaintop removal mining and valley filling on the occupancy and abundance of stream salamanders. Journal of Applied Ecology)

A data frame with 644 observations on the following 10 variables:

- *site*: name of a location where repeated samples were taken
- *mined*: factor indicating whether the site was affected by mountain top removal coal mining
- *cover*: amount of cover objects in the stream (scaled)
- *sample*: repeated sample
- *DOP*: Days since precipitation (scaled)
- *Wtemp*: water temperature (scaled)
- *DOY*: day of year (scaled)
- *spp*: abbreviated species name, possibly also life stage
- *count*: number of salamanders observed

Question of interest: What is the association between mining and the number of salamanders found at sites (in terms of occupancy and conditional abundance)?

```
head(Salamanders, 3)
```

```
##   site mined      cover sample      DOP      Wtemp      DOY spp count
## 1 VF-1   yes -1.4423172      1 -0.5956834 -1.22937861 -1.497003  GP     0
## 2 VF-2   yes  0.2984104      1 -0.5956834  0.08476529 -1.497003  GP     0
## 3 VF-3   yes  0.3978806      1 -1.1913668  1.01417627 -1.294467  GP     0
```

```
table(Salamanders$count)
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  36
## 387  79  61  30  29  17  10  12   6   3   2   3   1   2   1   1
```

Poisson zero heavy model

We can use **glmmTMB** to fit ZI- or hurdle models for longitudinal data. Note that the zero part is not modeled longitudinally.

```
fit.zip <- glmmTMB(count ~ mined + (1|site),
  zi=~mined,
  family=poisson,
  data=Salamanders)

summary(fit.zip)
```

```
## Family: poisson ( log )
## Formula:      count ~ mined + (1 | site)
## Zero inflation:      ~mined
## Data: Salamanders
##
##      AIC      BIC    logLik deviance df.resid
##  1908.5   1930.8   -949.2   1898.5     639
##
## Random effects:
##
## Conditional model:
##   Groups Name      Variance Std.Dev.
##   site (Intercept) 0.07843  0.28
## Number of obs: 644, groups: site, 23
##
## Conditional model:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.0879    0.2329   0.377   0.706
## minedno      1.1419    0.2462   4.639 3.5e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Zero-inflation model:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.1393    0.2351   4.846 1.26e-06 ***
## minedno      -1.7361    0.2620  -6.626 3.44e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The expected mean count of salamanders is $\exp(1.14)=3.13$ times in the sites with no mining than those sites with mining, for sites with non-zero salamanders. The odds of probability of zero structural counts for sites with no mining is $\exp(-1.74)=0.18$ times the odds of probability of zero structural counts for sites with mining.

Comparison of ZIP versus Poisson

```
fit.p <- glmmTMB(count ~ mined + (1|site),
  zi=~0,
  family=poisson, data=Salamanders)

summary(fit.p)

## Family: poisson ( log )
## Formula:      count ~ mined + (1 | site)
## Data: Salamanders
##
##      AIC      BIC    logLik deviance df.resid
##  2215.7   2229.1  -1104.8   2209.7     641
##
## Random effects:
##
## Conditional model:
##   Groups Name      Variance Std.Dev.
```

```
## site (Intercept) 0.3316 0.5759
## Number of obs: 644, groups: site, 23
##
## Conditional model:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.5053 0.2230 -6.749 1.49e-11 ***
## minedno 2.2644 0.2803 8.080 6.49e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## likelihood ratio test for whether or not the zero-inflation is needed
anova(fit.zip, fit.p)

## Data: Salamanders
## Models:
## fit.p: count ~ mined + (1 | site), zi=~0, disp=~1
## fit.zip: count ~ mined + (1 | site), zi=~mined, disp=~1
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## fit.p 3 2215.7 2229.1 -1104.85 2209.7
## fit.zip 5 1908.5 1930.8 -949.23 1898.5 311.23 2 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hurdle Poisson model

```
fit.hurdle <- glmmTMB(count ~ mined + (1|site),
  zi=~ mined,
  family=truncated_poisson,
  data=Salamanders)

## summary(fit.hurdle)
```