

Stat 534 Homework 1

Marty Simonson

October 3, 2019

1.) *Alseis blackiana*

Given a data file on abundance of *Alseis blackiana* in two habitats (OldHigh and OldSlope). Assume that the counts follow a Negative Binomial distribution, and paramaterize in terms of the mean μ and overdispersion parameter θ .

a)

Ignore habitat and find the MLE of mean and θ , the overdispersion parameter.

```
## [1] 20.40005
```

```
## [1] 0.3073573
```

The overall numerical optimization of the maximum likelihood parameters μ and θ are 20.400 and 0.307, respectively.

b)

Fit a model that allows habitat-specific means and shared overdispersion. Find the mle's of the difference in means (as OldHigh - OldSlope) and the overdispersion.

```
## [1] 17.87489
```

```
## [1] 0.1406785
```

The MLE for the difference in means between OldHigh - OldSlope is 17.875 and the MLE of the overdispersion parameter theta is 0.14

c)

Use the results from (b) to construct a normal approximation 95% confidence interval for the difference in means.

```
##      lower      upper
```

```
## 8.255383 27.494390
```

The 95% normal approximation for the difference in means between habitat is between 8.255 to 27.494.

d)

Use results from this model (and perhaps other models) to construct a likelihood ratio test **LRT** of the $H_0 : \mu_{OldHigh} = \mu_{OldSlope}$. Report the value of the test statistic and and p-value.

```
# reduced - full model
LRT1<- -2*(one.a$value - fit.b$value)
LRT1
```

```
## [1] 13.50746
```

```
pval.lrt1<-pchisq(LRT1, df=1, lower.tail=FALSE)
pval.lrt1
```

```
## [1] 0.0002376176
```

The test statistic is 13.507 with an associated p-value of 0.000238, leading us to the conclusion that there is evidence to reject the null hypothesis that the mean counts between habitats are different.

e)

Fit a model (or models) that allows habitat-specific means and habitat specific overdispersion. Find the MLE's of the two means and the two overdispersion parameters.

```
## [1] 0.09981814
```

```
## [1] 0.1805232
```

```
## [1] 13.24947
```

```
## [1] 31.124
```

The MLE's for the mean in habitats Old Slope and Old High are 13.249 and 31.127, respectively. The MLE's for overdispersion in Old Slope and Old High habitats are 0.0997 and 0.181, respectively.

f)

Use results from this model (and perhaps other models) to construct a likelihood ratio test of $H_0 : \theta_{OldHigh} = \theta_{OldSlope}$. Report the value of your test statistic and a p-value.

```
# -2 * (reduced - full model)
LRT2<- -2*(fit.b$value - fit.e$value)
LRT2
```

```
## [1] 0.4610358
```

```
pval.lrt2<-pchisq(LRT2, df=1, lower.tail=FALSE)
pval.lrt2
```

```
## [1] 0.4971403
```

The test statistic is 0.461 with an associated p-value of 0.497, suggesting that there is no evidence to support the alternative hypothesis that there is a difference in θ between the two habitats.

3 Missouri Fish Communities

- (a) Is there an interaction between site and habitat, i.e., is the difference between the two habitats similar (or not) at the three sites?
- (b) If there is an interaction, which species are significant contributors to that interaction?

Answer:

Methods

Missouri fish sampling was conducted where two habitats (main channel and chute) were sampled at each of three sites using five different gears. Species of fish found in fewer than 10% of samples have been omitted from the data. The objective is to test whether or not species abundance is affected by an interaction between site and habitat. If an interaction exists, we will determine which species are significant contributors to the observed interaction of site and habitat.

I assume that since we are given count data that our models are overdispersed and will create a model under the negative binomial distribution. I further assume that each species is independent.

The full model includes additive effects of habitat, site, and gear as well as interactions of site:gear, gear:habitat, and our parameter of interest site:habitat. The reduced (nested) model included the same individual effects of habitat, site, and gear but also had interactions of habitat:gear and gear:site.

Both models were implemented using the R packages “mvabund” and “vegan”, particularly the function *manyglm()* that is used to calculate parameter estimates of generalized linear models fitted to each of many variable simultaneously. A log-link function was used with the default negative binomial distribution. Significance for the interaction between site and habitat was tested using in the full model using a bootstrap analysis of variance function set to 999 iterations (known as PIT-trap resampling).

Fish species that were significant contributors to any observed interaction between site and habitat were determined by plotting the deviance, or likelihood ratio test, between our full and reduced models for each species. The threshold for significance was determined first by a Bonferroni-corrected chi-square test and then by a traditional chi-square test, each with $\alpha = 0.05$ and two degrees of freedom.

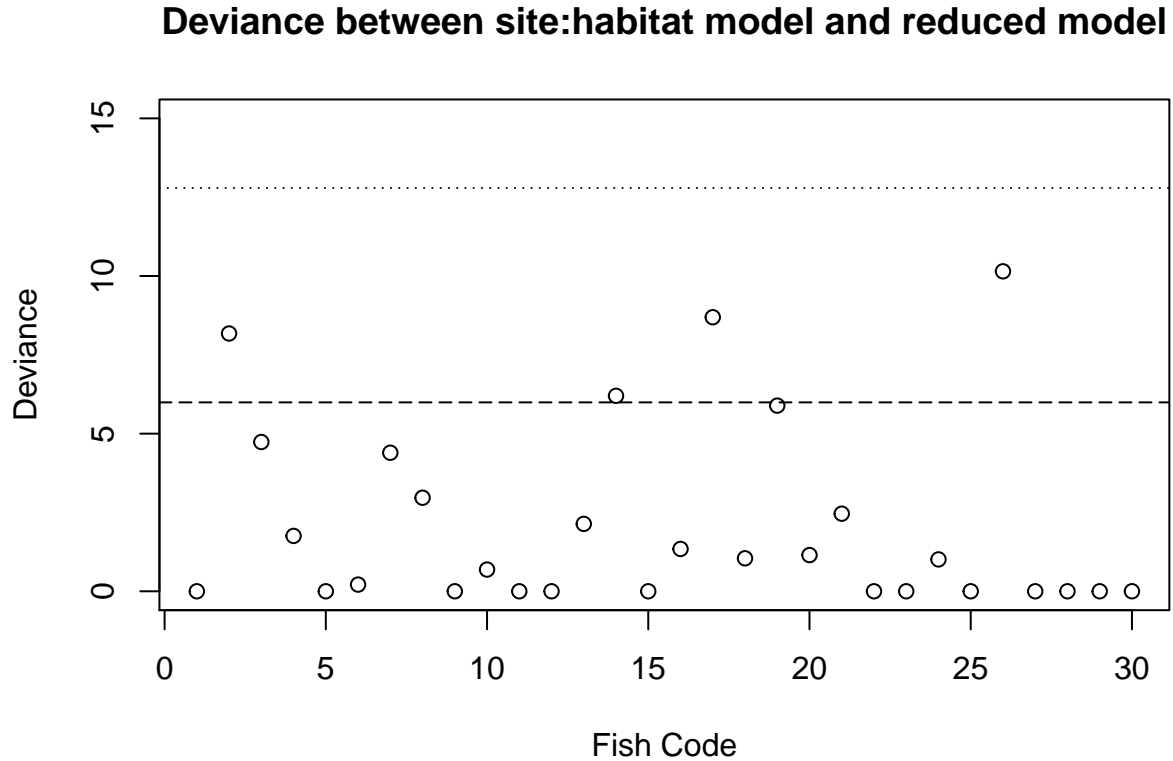
Results

The ANOVA for the full model showed that there was some evidence for an interaction between site and habitat (p-value ~ 0.05). The model for ‘gear’ was highly significant (p-value ~ 0.01) and models with ‘site’ and ‘site:gear’ were observed to be significant given the data.

```
## Time elapsed: 0 hr 4 min 5 sec

## Analysis of Deviance Table
##
## Model: manyglm(formula = fish.mva ~ habitat + site + gear + site * gear +
## Model:      gear * habitat + site * habitat, data = fish.env)
##
## Multivariate test:
##           Res.Df Df.diff    Dev Pr(>Dev)
## (Intercept)      54
## habitat          53      1  45.0   0.181
## site            51      2 110.9   0.018 *
## gear            47      4 582.5   0.001 ***
## site:gear        39      8 101.4   0.035 *
## habitat:gear     35      4  34.0   0.804
## habitat:site     33      2  63.0   0.039 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Arguments:
## Test statistics calculated assuming uncorrelated response (for faster computation)
## P-value calculated using 999 resampling iterations via PIT-trap resampling (to account for correlat
```

Plotting the deviance between models yielded zero species that significantly contributed to the observed interaction between site and habitat under the Bonferroni-adjusted chi-square test (dotted line). However, the traditional chi-square threshold (dashed line) determined that there were four species that were likely significant contributors to the observed interaction between sample site and sample habitat.



The four species that were common carp *Cyprinus carpio*, channel catfish *Ictalurus punctatus*, red shiner *Cyprinella lutrensis*, and silver chub *Machrybopsis soteriana*. The first two species are large-bodied bottom feeders, and may use habitat differently based on water current (chute vs. main channel). The other two species are small-bodied individuals and may not have the strength to withstand current in the main channel.