Harbor seals analysis

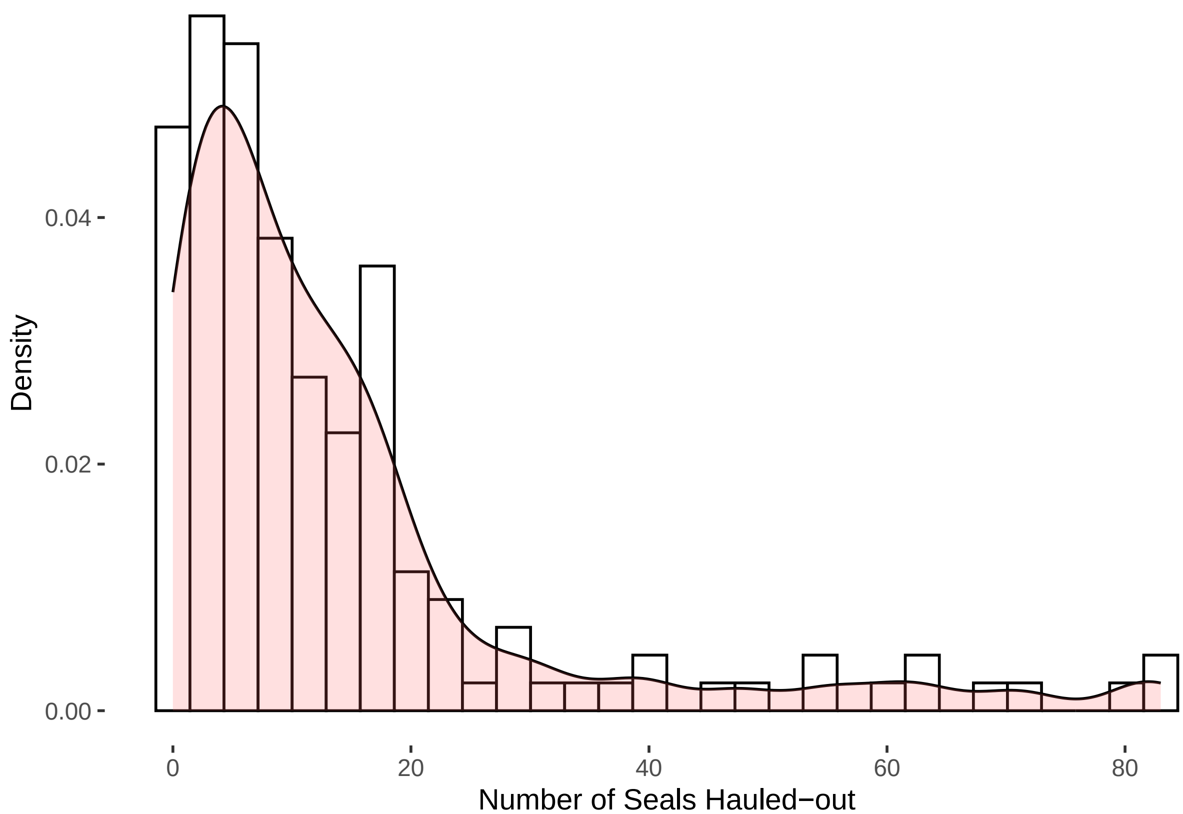
Kyra Bankhead

2022-12-14

In this markdown I will:

1. Find the best distribution to use for GLMs.
2. Run GLMS and AICs to find the most appropriate model and predictors.
3. Create visualization graphs for each site.

Check for Appropriate Distribution



Negative binomial or poisson would be the best fit. Let's check if the mean is equal to the variance and if there is zero inflation.

```
full.data<-read.csv("../data/full.data.csv")
require(performance)
```

```
## Loading required package: performance
```

```r
var(full.data$seals)
```

```
## [1] 299.2771
```

```r
mean(full.data$seals)
```

```
## [1] 14.12258
```

```r
# Make sure overdispersion is detected with full model
mod<- glm(seals ~ site*noise + month + tide + time, data = full.data, family = "poisson")
check_overdispersion(mod)
```

```
## # Overdispersion test
##
##        dispersion ratio =   6.737
##   Pearson's Chi-Squared = 997.052
##                 p-value = < 0.001
```

```
## Overdispersion detected.
```

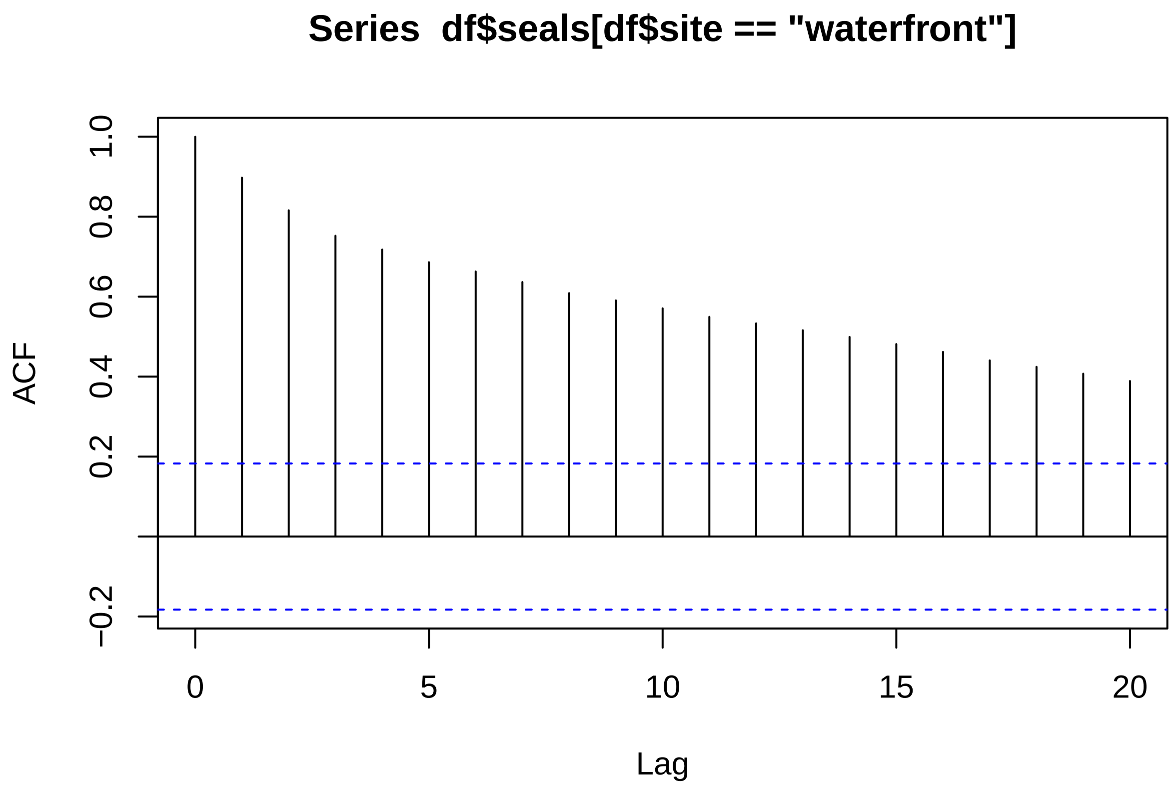
```r
# Check for zero inflation
check_zeroinflation(mod)
```

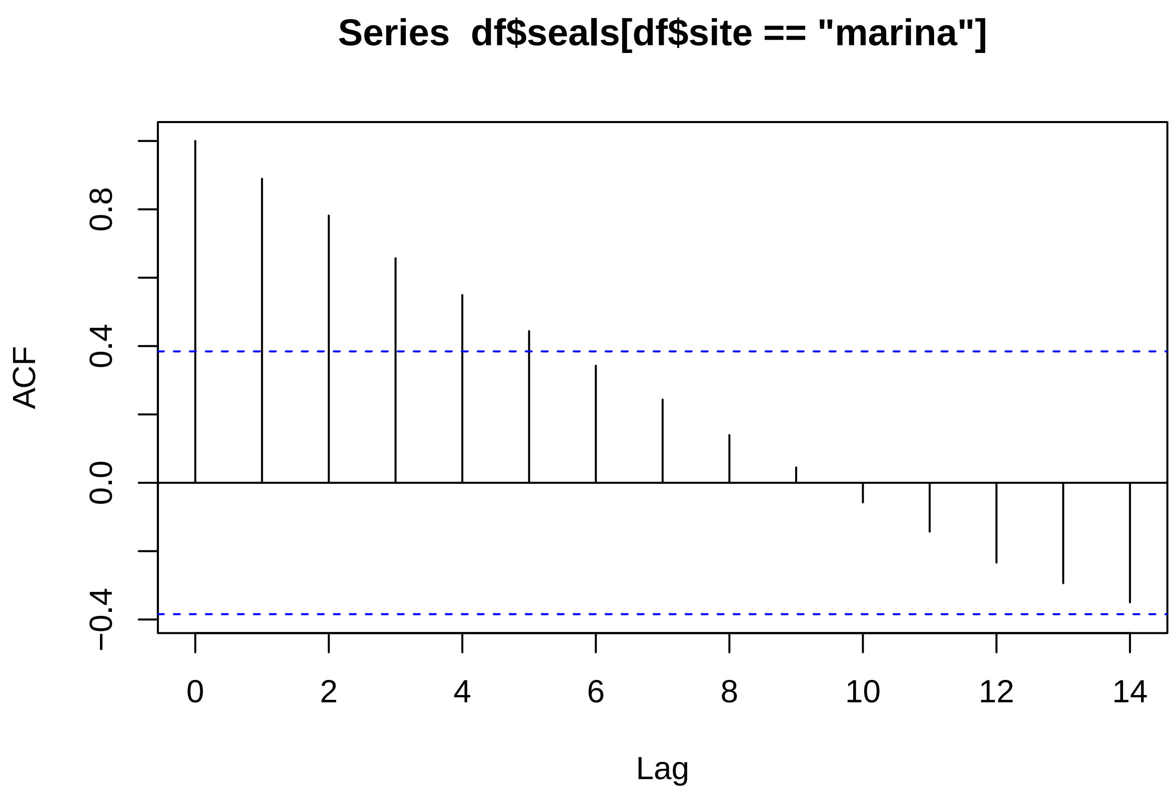
```
## # Check for zero-inflation
##
##    Observed zeros: 14
##   Predicted zeros: 0
##             Ratio: 0.00
```

```
## Model is underfitting zeros (probable zero-inflation).
```

The variance is way higher than the mean which was confirmed by the Pearson chi-squared test. Additionally there was zero-inflation present in the model. Therefore the Poisson distribution is not applicable here and we will continue with the negative binomial distribution.

Check for autocorrelation





The x-axis corresponds to the different lags of the residuals. The y-axis shows the correlation of each lag. Finally, the dashed blue line represents the significance level. Both sites show high autocorrelation. However, there is no way to incorporate overdispersion into a model that incorporates autocorrelation such as GEEs. Therefore we will stick with GLMs.

Create GLMs and find best model with AICc

To test whether noise affects the number of seals hauled-out by site, I will insert an interaction between noise level and site.

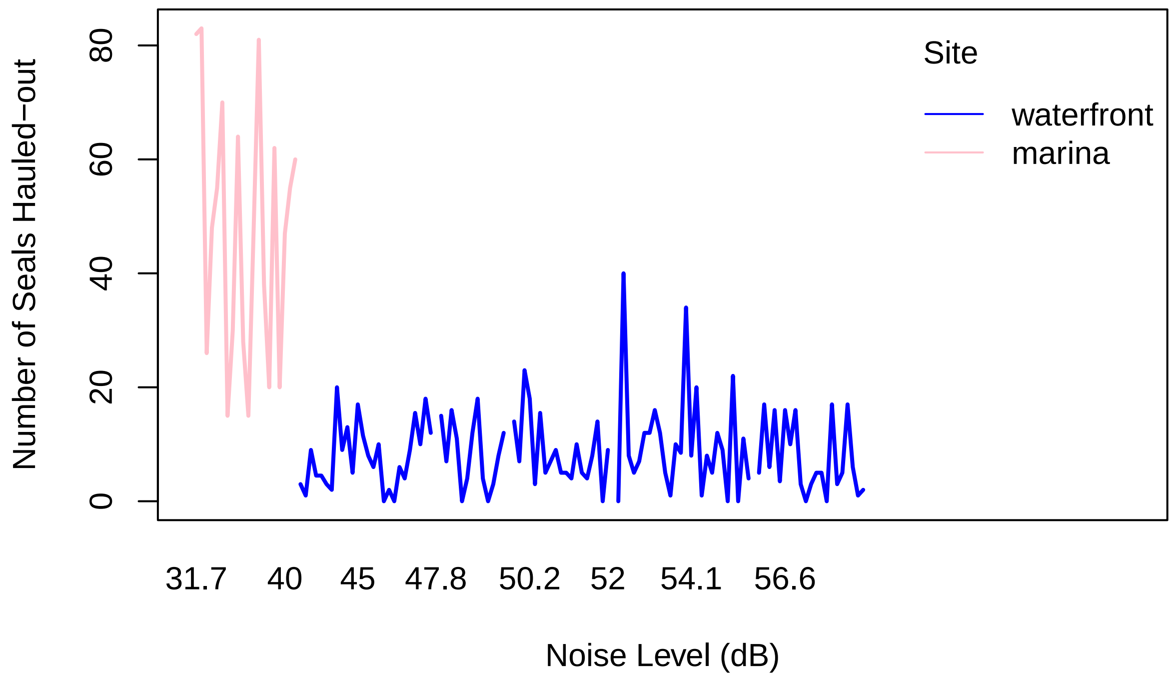
```
##
## Model selection based on AICc:
##
##                                         K    AICc Delta_AICc AICcWt Cum.Wt
## seals ~ site*noise + month + time       7 1060.03       0.00   0.57   0.57
## seals ~ site*noise + month + tide + time 8 1061.88       1.85   0.23   0.80
## seals ~ site*noise + month              6 1062.17       2.14   0.20   1.00
## seals ~ 1                               2 1143.70      83.67   0.00   1.00
##                                              LL
## seals ~ site*noise + month + time       -522.63
## seals ~ site*noise + month + tide + time -522.45
## seals ~ site*noise + month              -524.80
## seals ~ 1                               -569.81
```

Looks like the best model will contain month, noise, site and time as predictors. This is the summary of that model:

```
##
## Call:
## glm.nb(formula = seals ~ site * noise + month + time, data = full.data,
##     init.theta = 1.738579161, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9338  -0.8665  -0.1937   0.4408   2.4225
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)           8.31830    1.07277   7.754 8.90e-15 ***
## sitewaterfront       -3.62238    1.25922  -2.877  0.00402 **
## noise                -0.06368    0.02554  -2.493  0.01266 *
## month                -0.18974    0.04597  -4.127 3.67e-05 ***
## time                 -0.05820    0.02795  -2.082  0.03733 *
## sitewaterfront:noise  0.05995    0.02923   2.051  0.04028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.7386) family taken to be 1)
##
##     Null deviance: 319.08  on 154  degrees of freedom
## Residual deviance: 183.65  on 149  degrees of freedom
## AIC: 1059.3
##
## Number of Fisher Scoring iterations: 1
##
##
```

```
##                Theta:  1.739
##            Std. Err.:  0.250
##
##  2 x log-likelihood:  -1045.269
```

- Month and time are significant predictors for how many harbor seals haul-out.
- Site and noise are significant predictors for the number of harbor seals hauled-out. The effect of noise on the number of seals haul-out depends on what site they are located in.



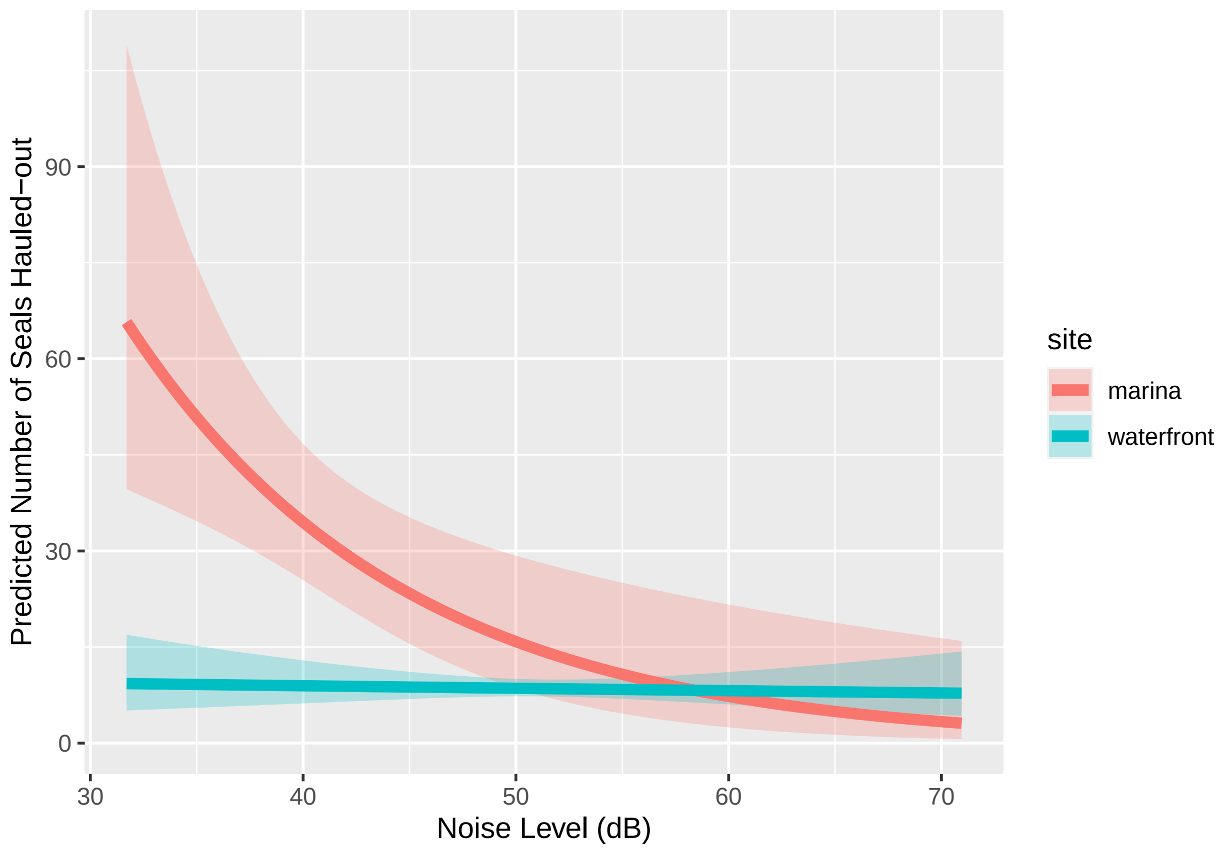
Use model output to predict response variable

```
##   noise       site      phat
## 1 49.07     marina 17.034396
## 2 49.07 waterfront  8.595981
```

In the output above, we see that the predicted number of seals hauled-out for the Marina is about 17.03, holding noise level at its mean. The predicted number of events for the Waterfront is lower at 8.60.

Below we will obtain the mean predicted number of seals hauled-out for values of noise across its entire range for each site and graph these.

```
##   noise       site      phat
## 1 49.07     marina 17.034396
## 2 49.07 waterfront  8.595981
```



The graph shows the log linear model of the expected seals counts across the range of noise levels, for each site along with 95 percent confidence intervals. Note that what is plotted are the expected values, not the log of the expected values.