

# Modeling Poisson Counts

## Poisson regression – Stat 230

In this class we'll explore how to use Poisson regression models to *describe counts rather than rates*. The toolkit (inference, diagnostics, etc.) for Poisson regression remains the same, but there are a few key differences when fitting the model.

### Data overview

Male double-crested cormorants use advertising behavior to attract females for breeding. The **Cormorants** data set gives some results from a study by Meagan Mc Rae (2015) on counts of advertising males observed two or three times in a week at six stations in a tree-nesting colony for an entire breeding season. The number of advertising birds was counted and these observations were classified by characteristics of the trees and nests. The goal was to determine how this behavior varies temporally over the season and spatially over observation stations, as well as with characteristics of nesting sites. You can load the **Cormorants** data via

```
cormorants <- read.csv("http://aloy.rbind.io/data/Cormorants.csv")
```

In this activity you will use **count** as the response variable. Below is a quick overview of the other variables:

Variable	Description
<b>category</b>	Time of season, divided into 3 categories based on breeding chronology, an ordered factor with levels <b>Pre</b> < <b>Incubation</b> < <b>Chicks Present</b>
<b>week</b>	Week of the season
<b>station</b>	Station of observations on two different peninsulas in a park, a factor with levels <b>B1</b> , <b>B2</b> , <b>C1</b> , <b>C2</b> , <b>C3</b> , <b>C4</b>
<b>nest</b>	Type of nest, an ordered factor with levels <b>no</b> < <b>partial</b> < <b>full</b>
<b>height</b>	Relative height of bird in the tree, an ordered factor with levels <b>low</b> < <b>mid</b> < <b>high</b>
<b>density</b>	Number of other nests in the tree, an ordered factor with levels <b>zero</b> < <b>few</b> < <b>moderate</b> < <b>high</b>
<b>tree_health</b>	Health of the tree the bird is advertising in, a factor with levels <b>dead</b> and <b>healthy</b>
<b>count</b>	Number of birds advertising

## Fitting Poisson regression models to count data

To fit a Poisson regression model for counts (rather than rates) we still use the `glm()` command and specify `family = poisson`, but now we can omit the `offset` argument.

1. Fit a Poisson regression model with `week` and `station` as the two predictor variables. Report the fitted model equation.
2. Interpret each exponentiated slope estimate in context. Hint: You will need to think about what level of `station` is the baseline (i.e., the one without an indicator variable).
3. Next, let's plot the fitted model. To do this, we can use effect plots. Once you have loaded the `{effects}` package, plug your fitted model into the blank and run the code.

```
plot(allEffects(___), main = "", rescale.axis = FALSE)
```

4. Rerun the code for the effects plot, but this time set `rescale.axis = TRUE`. How does the plot change? What do you notice about the scale of the y-axis? Does this make sense?

## Conducting inference

5. Conduct an appropriate hypothesis test to determine if `station` is an important predictor. Report the hypotheses, the test statistic, the p-value, and your conclusion in context.
6. Construct an 89% Wald confidence interval for the `week` predictor.

## Model checking

7. To construct deviance residual plots we can use the same `residualPlots()` command from the `{car}` package. Do you see any signs of non-linearity? (Note: in order for the residual plot for `station` to show up, you should wrap `specify factor(station)` when you fit the model. I have no clue why this is needed, but apparently it is.)
8. Explore whether the effect of `week` should be linear in the model. To do this, consider whether a polynomial for `week` substantially improves the model. Report your findings.

## Extending the model to more predictors

Rather than focusing only on `week` and `station` as predictors, let's expand our focus to include all possible predictors.

9. Fit a Poisson regression model with all of the predictors included as main effects. If you found a polynomial for `week` was appropriate, then you should include that term here as well. Report the fitted model equation.
10. Create an effect plot to help you decipher your model results.