Zero-Inflated Negative Binomial regression

- We are going to use the variables: **child** and **camper** to model the count in the part of negative binomial model and the variable **persons** in the logit part of the model.
- We use the **pscl** to run a zero-inflated negative binomial regression.
- We begin by estimating the model (called m1) with the variables of interest.

```
m1 <- zeroinfl(count ~ child + camper | persons,
  data = fishing, dist = "negbin",
  EM = TRUE)

summary(m1)
```

```
## Call:
## zeroinfl(formula = count ~ child + camper | persons,
##     data = fishing,
##     dist = "negbin", EM = TRUE)
##
## Pearson residuals:
##    Min     1Q Median     3Q    Max
## -0.586 -0.462 -0.389 -0.197 18.013
```

- ▶ Below the model call, you will find a block of output
  containing negative binomial regression coefficients for each of
  the variables along with standard errors, z-scores, and p-values
  for the coefficients.

- ▶ A second block follows that corresponds to the inflation
  model. This includes logit coefficients for predicting excess
  zeros along with their standard errors, z-scores, and p-values.

```
## Count model coefficients (negbin with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)     1.371      0.256    5.35  8.6e-08 ***
## child          -1.515      0.196   -7.75  9.4e-15 ***
## camper1         0.879      0.269    3.26   0.0011 **
## Log(theta)     -0.985      0.176   -5.60  2.1e-08 ***
```

```
## Zero-inflation model coefficients (binomial with logit l
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.603      0.836    1.92    0.055 .
## persons       -1.666      0.679   -2.45    0.014 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Theta = 0.373
## Number of iterations in BFGS optimization: 2
## Log-likelihood: -433 on 6 Df
```

Tests of Significance

- ▶ All of the predictors in both the count and inflation portions of the model are statistically significant.
- ▶ This model will fit the data significantly better than the null model, i.e., the intercept-only model.
- ▶ To show that this is the case, we could compare with the current model to a null model without predictors using chi-squared test on the difference of log likelihoods.