PART 4 :

Zero-Inflated Negative Binomial Regression

Zero-Inflated Negative Binomial regression

- Zero-inflated negative binomial regression is for modeling count variables with excessive zeros and it is usually for overdispersed count outcome variables.
- Furthermore, theory suggests that the excess zeros are generated by a separate process from the count values and that the excess zeros can be modeled independently.

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.

Zero-Inflated Negative Binomial regression

```
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
```

Zero-Inflated Negative Binomial regression

- 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.

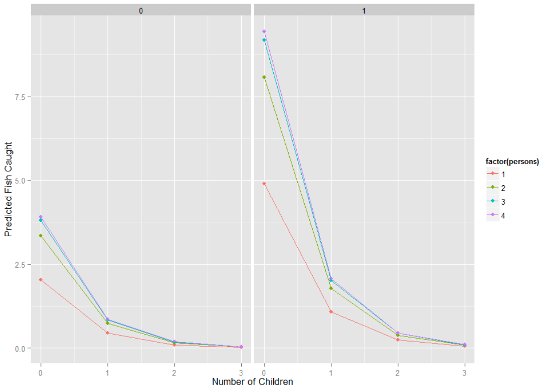
```
## Count model coefficients (negbin with log link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.371      0.256    5.35  8.6e-08 ***
## child         -1.515      0.196   -7.75  9.4e-15 ***
## camper1        0.879      0.269    3.26   0.0011 **
## Log(theta)    -0.985  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

- The predictors child and camper in the part of the negative binomial regression model predicting number of fish caught (count) are both significant predictors.
- The predictor person in the part of the logit model predicting excessive zeros is statistically significant.
- For these data, the expected change in log(count) for a one-unit increase in child is -1.515255 holding other variables constant.
- A camper (camper $= 1$) has an expected log(count) of 0.879051 higher than that of a non-camper (camper $= 0$) holding other variables constant.

Tests of Significance

- All of the predictors in both the count and inflation portions of the model are statistically significant.


0
1
Predicted Fish Caught
7.5
5.0
2.5
0.0
0
1
2
3
Number of Children
factor(persons)
1
2
3
4

- The log odds of being an excessive zero would decrease by 1.67 for every additional person in the group.
- In other words, the more people in the group the less likely that the zero would be due to not gone fishing.
- Put plainly, the larger the group the person was in, the more likely that the person went fishing.
- The Vuong test suggests that the zero-inflated negative binomial model is a significant improvement over a standard negative binomial model.