Zero-inflated Regression models - Summary

- Zero-inflated models attempt to account for excess zeros.
- In other words, two kinds of zeros are thought to exist in the data, "**true zeros**" and "**excess zeros**".

Zero-inflated Regression models

Two Distinct Processes

- The two parts of the a zero-inflated model are a binary model, usually a logit model to model which of the two processes the zero outcome is associated with and a count model, in this case, a negative binomial model, to model the count process.
- In other words, the excess zeros are generated by a separate process from the count values and that the excess zeros can be modelled independently.
- Zero-inflated models estimate two equations simultaneously, one for the count model and one for the excess zeros.
- The expected count is expressed as a combination of the two processes.

The `zeroinfl()` function

- In `R`, zero-inflated count data models can be fitted with the `zeroinfl()` function from the **pscl** package.
- Both the fitting function interface and the returned model objects of class "`zeroinfl`" are modelled after the corresponding GLM functionality in `R`.

The arguments of zeroinfl() are given by

```
zeroinfl(formula, data, subset, na.action,
        weights, offset,
        dist = "poisson", link = "logit",
        control = zeroinfl.control(...),
        model = TRUE, y = TRUE, x = FALSE, ...)
```

Zero-inflated Regression models

Fishing Data Set

- We have data on 250 groups that went to a park.
- Each group was questioned about how many fish they caught (**count**), how many children were in the group (**child**), how many people were in the group (**persons**), and whether or not they brought a camper to the park (**camper**).
- In addition to predicting the number of fish caught, there is interest in predicting the existence of excess zeros, i.e., the probability that a group caught zero fish.
- We will use the variables child, persons, and camper in our model.

Zero-inflated Regression models

Fishing Data Set

- In addition to predicting the number of fish caught, there is interest in predicting the existence of excess zeros, i.e., the probability that a group caught zero fish.
- We will use the variables child, persons, and camper in our model.

```
> head(fish)
nofish livebait camper persons child          xb
1      1        0      0       1     0 -0.8963146
2      0        1      1       1     0 -0.5583450
3      0        1      0       1     0 -0.4017310
4      0        1      1       2     1 -0.9562981
5      0        1      0       1     0  0.4368910
6      0        1      1       4     2  1.3944855
zg count
1  3.0504048     0
2  1.7461489     0
3  0.2799389     0
4 -0.6015257     0
5  0.5277091     1
6 -0.7075348     0
```

What is a Zero-Inflated Model?

The Fishing Example

- A zero-inflated model assumes that zero outcome is due to two different processes.
- For instance, in the example of fishing presented here, the two processes are that a subject has gone 1. *fishing* vs. 2. *not fishing*.
- If not gone fishing, the only outcome possible is zero.
- If gone fishing, it is then a count process.

$$E(\text{no. fish caught} = k)$$

$$= P(\text{not fishing}) \times 0 + P(\text{fishing}) \times E(y = k|\text{fishing})$$