Homework #5

Dustin Leatherman

2/24/2020

The California state wildlife biologists want to model how many fish are being caught by fishermen at a state park. Visitors are asked how long they stayed, how many people were in the group, were there children in the group and how many fish were caught. Some visitors do not fish, but there is no data on whether a person fished or not. Some visitors who did fish did not catch any fish so there are excess zeros in the data because of the people that did not fish.

There were 250 groups that went to a park. each group ws questioned about how many fish they acught (count), how many children were in the group (child), how many people wer ein the group (persons), and whether or not they brought a camper to the part (camper).

a

> Use Poisson regression to model how many fish they caught (count) using person and camper as independent variables Interpret the effect of the significant independent variables.

```r
fish.m.poisson <- glm(count ~ persons + camper, family = "poisson", data = fish)

tidy(fish.m.poisson) %>%
  mutate_at(c("estimate", "std.error"), exp) %>%
  kable(
    caption = "Estimated Incident Effects for Predictors"
  ) %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

Table 1: Estimated Incident Effects for Predictors

term	estimate	std.error	statistic	p.value
(Intercept)	0.1064757	1.177095	-13.73719	0
persons	2.3708755	1.041667	21.14695	0
camper	3.2986258	1.092882	13.43767	0

```r
confint(fish.m.poisson) %>%
  exp %>%
  kable(
    caption = "Estimated Incident Effects C.I"
  ) %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

```
## Waiting for profiling to be done...
```

There is convincing evidence that the number of group members and whether or not they had a camper significantly affect the number of fish caught (two-tailed t-test).

It is estimated that for each additional person in the group, the number of fish caught increases 2.371 times. With 95% confidence, the number of fish caught increases by 2.191 to 2.571 times for each additional person in the group.

Table 2: Estimated Incident Effects C.I

	2.5 %	97.5 %
(Intercept)	0.0769059	0.145750
persons	2.1910136	2.571389
camper	2.7799124	3.938604

It is estimated that if a group brought a camper, the number of fish caught increases by 3.299 times. With 95% confidence, the number of fish caught increases by 2.78 to 3.939 times when the group has a camper.

b

Use Negative Binomial regression to model how many fish they caught (count) using person and camper as independent variables Interpret the effect of the significant independent variables.

```
fish.m.negbin <- glm.nb(count ~ persons + camper, data = fish)

tidy(fish.m.negbin) %>%
  mutate_at(c("estimate", "std.error"), exp) %>%
  kable(
    caption = "Estimated Incident Effects for Predictors"
  ) %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

Table 3: Estimated Incident Effects for Predictors

term	estimate	std.error	statistic	p.value
(Intercept)	0.1923451	1.499320	-4.070162	0.0000470
persons	2.1176088	1.132047	6.049347	0.0000000
camper	2.3353708	1.321513	3.042466	0.0023465

```
confint(fish.m.negbin) %>%
  exp %>%
  kable(
    caption = "Estimated Incident Effects C.I"
  ) %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

```
## Waiting for profiling to be done...
```

Table 4: Estimated Incident Effects C.I

	2.5 %	97.5 %
(Intercept)	0.0912592	0.4128483
persons	1.6619379	2.7067085
camper	1.3376712	4.0220205

There is convincing evidence that the number of group members and whether or not they had a camper significantly affect the number of fish caught (two-tailed t-test).

It is estimated that for each additional person in the group, the number of fish caught increases 2.12 times.

With 95% confidence, the number of fish caught increases by 1.662 to 2.707 times for each additional person in the group.

It is estimated that if a group brought a camper, the number of fish caught increases by 2.34 times. With 95% confidence, the number of fish caught increases by 1.338 to 4.022 times when the group has a camper.

C

> Use Zero-Inflated Poisson regression to model how many fish they caught (count) using person and camper as independent variables. Interpret the effect of the significant independent variables.

```
fish.m.zip <- zeroinfl(count ~ persons + camper, data = fish)

summary(fish.m.zip)$coefficients$count %>%
  as_tibble %>%
  mutate_at(c("Estimate", "Std. Error"), exp) %>%
  kable(
    caption = "Estimated Incident Effects for Predictors"
  ) %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

Table 5: Estimated Incident Effects for Predictors

Estimate	Std. Error	z value	Pr(>|z|)
0.5180578	1.180990	-3.953455	7.7e-05
2.0651249	1.043906	16.876987	0.0e+00
1.8069029	1.095701	6.473205	0.0e+00

```
confint(fish.m.zip) %>%
  exp %>%
  kable(
    caption = "Estimated Incident Effects C.I"
  ) %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

Table 6: Estimated Incident Effects C.I

	2.5 %	97.5 %
count_(Intercept)	0.3739198	0.7177580
count_persons	1.8983263	2.2465794
count_camper	1.5105664	2.1613735
zero_(Intercept)	1.0147820	5.3926165
zero_persons	0.6970759	1.1535188
zero_camper	0.2942594	0.8851137

There is convincing evidence that the number of group members and whether or not they had a camper significantly affect the number of fish caught.

It is estimated that for each additional person in the group, the number of fish caught increases 2.07 times. With 95% confidence, the number of fish caught increases by 1.898 to 2.247 times for each additional person in the group.

Table 7: Estimated Incident Effects for Predictors

term	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.1970130	1.461657	-4.279793	0.0000187
persons	2.3393041	1.129392	6.984357	0.0000000
camper	1.7495173	1.355602	1.838447	0.0659966
Log(theta)	0.3042615	1.160472	-7.994974	0.0000000

Table 8: Estimated Incident Effects C.I

	2.5 %	97.5 %
count_(Intercept)	0.0936275	4.145590e-01
count_persons	1.8429471	2.969344e+00
count_camper	0.9637040	3.176090e+00
zero_(Intercept)	0.0000001	3.952936e+01
zero_persons	0.3932792	5.872441e+01
zero_camper	0.0000000	5.153534e+40

It is estimated that if a group brought a camper, the number of fish caught increases by 1.81 times. With 95% confidence, the number of fish caught increases by 1.511 to 2.161 times when the group has a camper.

d

> Use Zero-Inflated Negative Binomial regression to model how many fish they caught (count) using person and camper as independent variables. Interpret the effect of the significant independent variables.

```
fish.m.zinb <- zeroinfl(count ~ persons + camper, data = fish, dist = "negbin")

summary(fish.m.zinb)$coefficients$count %>%
  as_tibble(rownames = "term") %>%
  mutate_at(c("Estimate", "Std. Error"), exp) %>%
  kable(
    caption = "Estimated Incident Effects for Predictors"
  )
```

```
confint(fish.m.zinb) %>%
  exp %>%
  kable(
    caption = "Estimated Incident Effects C.I"
  )
```

Given the low p-value for Log(theta), there is convincing evidence that this model is **not** a Poisson Model. There is no evidence that the true model is a zero-inflated Negative Binomial model.

It is estimated that for each additional person in the group, the number of fish caught increases 2.34 times. With 95% confidence, the number of fish caught increases by 1.843 to 2.969 times for each additional person in the group.

It is estimated that if a group brought a camper, the number of fish caught increases by 1.75 times. With 95% confidence, the number of fish caught increases by 0.964 to 3.176 times when the group has a camper.

e

Which of the above models is best?

```
fish %>% dplyr::select(count) %>% filter(count == 0) %>% count/250
```

```
##       n
## 1 0.568
```

56.8% of the data contains 0 counts making a zero inflated model more likely.

```
pscl::odTest(fish.m.negbin)
```

```
## Likelihood ratio test of H0: Poisson, as restricted NB model:
## n.b., the distribution of the test-statistic under H0 is non-standard
## e.g., see help(odTest) for details/references
##
## Critical value of test statistic at the alpha= 0.05 level: 2.7055
## Chi-Square Test Statistic =  1645.6132 p-value = < 2.2e-16
```

Using the Liklihood Ratio Test, there is convincing evidence that the negative binomial model is a better fit than the poisson model (p.value $<$ 2.2e-16). This is futher confirmed since by the significance of Log(theta) in the Negative Binomial model. This narrows down our choice between the Negative Binomial and Zero-Inflated Negative Binomial model.

```
data.frame(
  model = c("ZINB", "ZIP", "NegBin", "Poisson"),
  aic = c(
    AIC(fish.m.zinb),
    AIC(fish.m.zip),
    AIC(fish.m.negbin),
    AIC(fish.m.poisson)
  )
) %>%
  arrange(aic) %>%
  kable %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

model	aic
ZINB	890.7922
NegBin	893.4482
ZIP	1807.1570
Poisson	2537.0615

Going by AIC, the Zero-Inflated Negaitve Binomial model is the best among the four. Despite the coefficients for the zero-generation component being non-significant, it appears that it's a slightly better fit. It would seem that the Zero-Inflated model is marginally better so there is not a strong argument against using the negative binomial model.