Zero inflated models

A new question

$$P(Y_i = y) = \begin{cases} e^{-\lambda_i}(1-\alpha_i) + \alpha_i & y = 0 \\ \dfrac{e^{-\lambda_i}\lambda_i^y}{y!}(1-\alpha_i) & y > 0 \end{cases}$$

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma_0 + \gamma_1 Age_i$$

$$\log(\lambda_i) = \beta_0 + \beta_1 EducationSome_i + \beta_2 EducationCollege_i + \beta_3 EducationAdv_i + \beta_4 Diabetes_i$$

New research question: for smokers, does the number of cigarettes smoked per day depend on age?

How would we answer this research question?

Inference

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma_0 + \gamma_1 Age_i$$

$$\log(\lambda_i) = \beta_0 + \beta_1 EducationSome_i + \beta_2 EducationCollege_i + \beta_3 EducationAdv_i + \beta_4 Diabetes_i + \beta_5 Age_i$$

Research question: for smokers, does the number of cigarettes smoked per day depend on age?

What are the null and alternative hypotheses?

$H_0: \beta_5 = 0$

$H_A: \beta_5 \neq 0$

Wald test

Poisson

```
m2 <- zeroinfl(cigsPerDay ~ education +
                 diabetes + age | age,
               data = heart_data)
summary(m2)
```

logistic

```
...
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.2063437  0.0342290  93.673  < 2e-16 ***
## education2  -0.0441195  0.0124809  -3.535 0.000408 ***
## education3  -0.0820388  0.0158604  -5.173 2.31e-07 ***
## education4  -0.0062453  0.0171640  -0.364 0.715965
## diabetes    -0.0241419  0.0386336  -0.625 0.532042
## age         -0.0056183  0.0006738  -8.338  < 2e-16 ***
...
```

$$Z = \frac{-0.0056}{0.00067} = -8.34 \qquad \text{p-value} \approx 0$$

Likelihood ratio test

2(Log likelihood (full) − 2 log likelihood (reduced))

= deviance(reduced) − deviance (full)

```
m2 <- zeroinfl(cigsPerDay ~ education +
                 diabetes + age | age,
               data = heart_data)
m2$loglik
```

```
## [1] -14023.42
```

```
m1 <- zeroinfl(cigsPerDay ~ education +
                 diabetes | age,
               data = heart_data)
m1$loglik
```

```
## [1] -14058.41
```

$G = 2(-14023.42 - (-14058.41)) = 70.18 \sim \chi^2_1$

pchisq(70.18, 1, lower.tail=F) ≈ 0