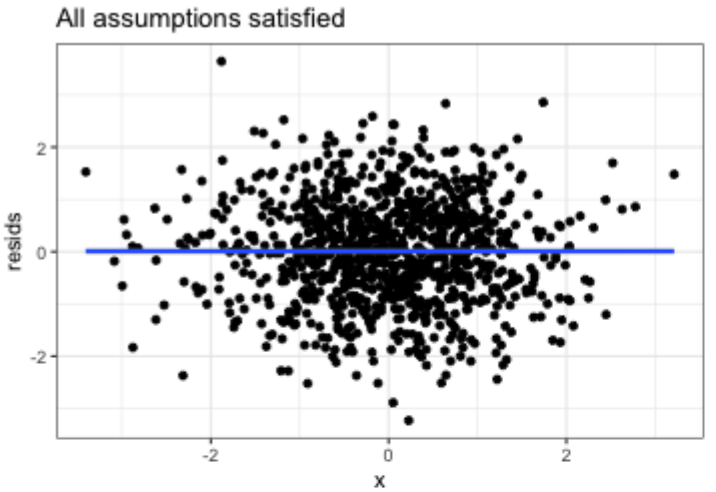
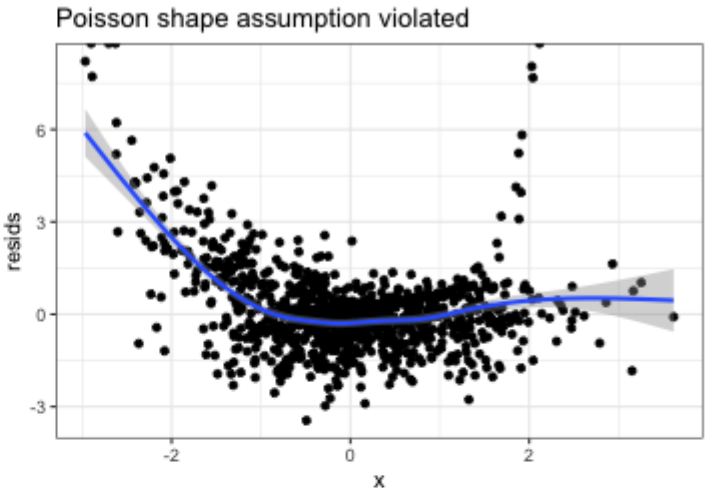
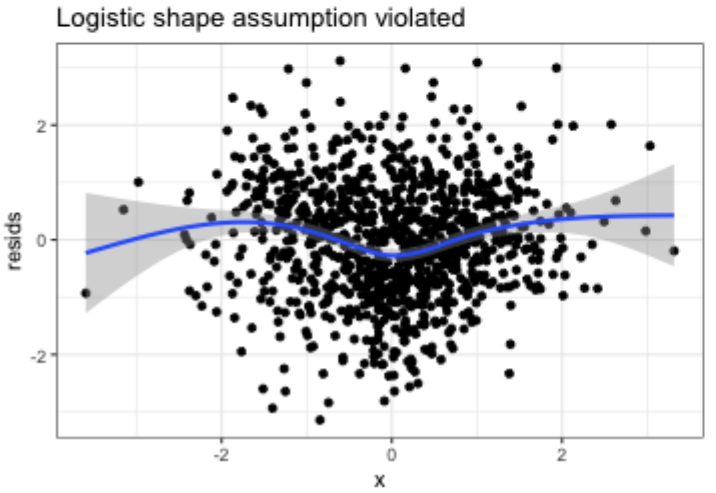
ZIP models

Recap: Assessing the shape assumption







Logistic component vs. Poisson component

Class activity

https://sta712-f22.github.io/class_activities/ca_lecture_33.html

Class activity

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma_0 + \gamma_1 EducationSome_i + \gamma_2 EducationCollege_i + \gamma_3 EducationAdv_i + \gamma_4 Diabetes_i + \gamma_5 Age_i$$

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

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

What are the null and alternative hypotheses?

Class activity

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma_0 + \gamma_1 EducationSome_i + \gamma_2 EducationCollege_i + \gamma_3 EducationAdv_i + \gamma_4 Diabetes_i + \gamma_5 Age_i$$

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

Research question: is there a relationship between age and whether someone is a smoker?

What are the null and alternative hypotheses?

Wald tests

> Research question: is there a relationship between age and whether someone is a smoker?

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

```
...
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.49673    0.20977 -11.902   <2e-16 ***
## education2  -0.06100    0.07840  -0.778   0.4366
## education3   0.17141    0.09362   1.831   0.0671 .
## education4   0.03547    0.10749   0.330   0.7414
## diabetes     0.26063    0.20854   1.250   0.2114
## age          0.05071    0.00395  12.838   <2e-16 ***
...
```

Class activity

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma_0 + \gamma_1 EducationSome_i + \gamma_2 EducationCollege_i + \gamma_3 EducationAdv_i + \gamma_4 Diabetes_i + \gamma_5 Age_i$$

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

Research question: Is there a relationship between education level and the number of cigarettes smoked?

What are the null and alternative hypotheses?

Likelihood ratio test

```
m1 <- zeroinfl(cigsPerDay ~ education + diabetes +
                 age | education + diabetes + age,
               data = heart_data)
m2 <- zeroinfl(cigsPerDay ~ education + diabetes
               | education + diabetes,
               data = heart_data)

2*(m1$loglik - m2$loglik)
```

```
## [1] 242.281
```

```
pchisq(242.281, df=6, lower.tail=F)
```

```
## [1] 1.828386e-49
```