Negative binomial regression

Recap: negative binomial regression

$$Y_i \sim NB(r,\ p_i)$$

$$\log(\mu_i) = \beta^T X_i$$

- $\mu_i = \dfrac{p_i r}{1 - p_i}$
- Note that r is the same for all i
- Note that just like in Poisson regression, we model the average count
  - Interpretation of βs is the same as in Poisson regression

In R

```r
library(MASS)
m2 <- glm.nb(cigsPerDay ~ male + age + education +
               diabetes + BMI, data = smokers)
```

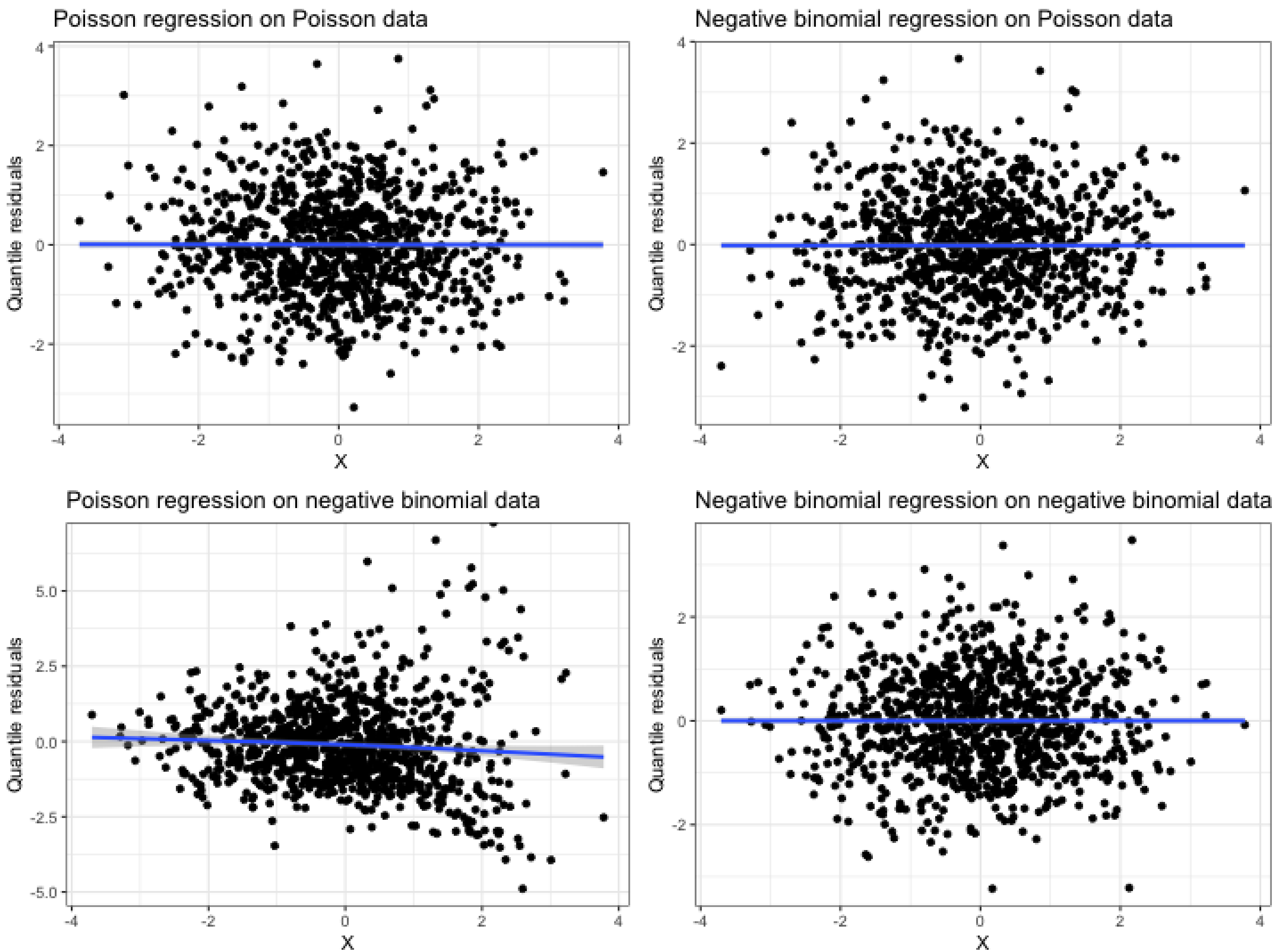
```
...
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.877771   0.123477  23.306  < 2e-16 ***
## male         0.459148   0.027641  16.611  < 2e-16 ***
## age         -0.007010   0.001731  -4.050 5.12e-05 ***
## education2   0.024518   0.032534   0.754    0.451
## education3   0.009252   0.040802   0.227    0.821
## education4  -0.027732   0.044825  -0.619    0.536
##
## (Dispersion parameter for Negative Binomial(3.2981) fami
...
```

$\hat{r} = 3.3$

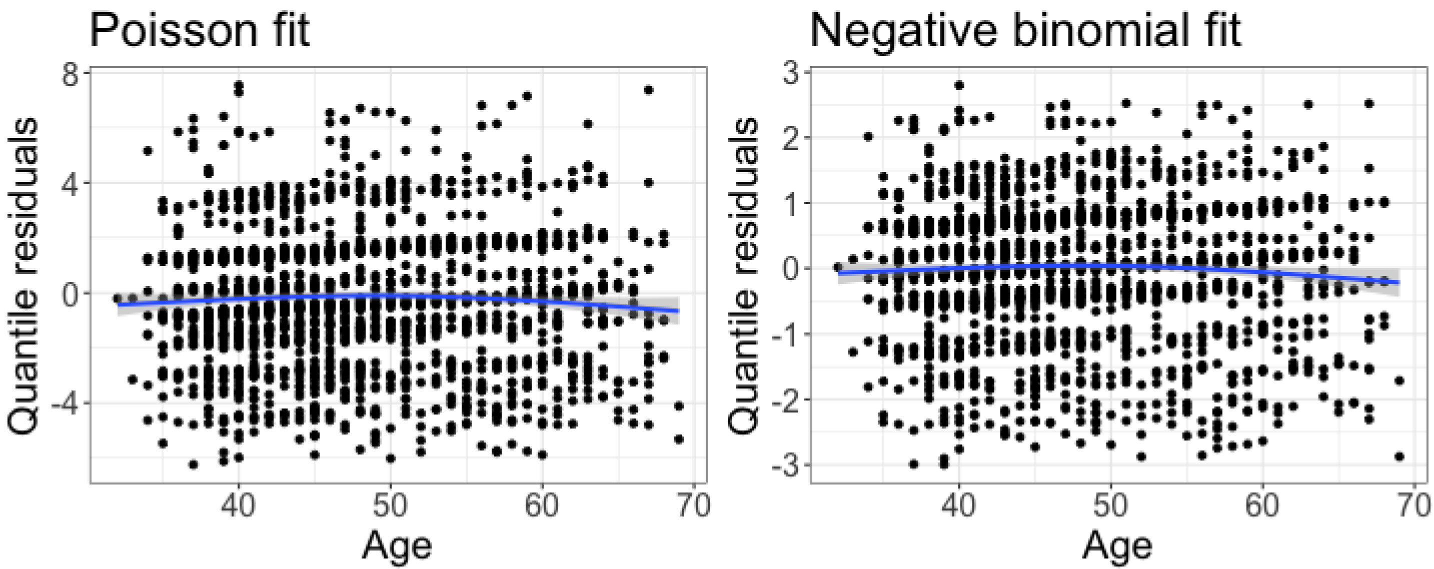
Class activity

https://sta712-f22.github.io/class_activities/ca_lecture_27.html

Class activity



Poisson vs. negative binomial fits



Inference with negative binomial models

```
...
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.877771   0.123477  23.306  < 2e-16 ***
## male         0.459148   0.027641  16.611  < 2e-16 ***
## age         -0.007010   0.001731  -4.050 5.12e-05 ***
## education2   0.024518   0.032534   0.754    0.451
## education3   0.009252   0.040802   0.227    0.821
## education4  -0.027732   0.044825  -0.619    0.536
## diabetes    -0.010124   0.099126  -0.102    0.919
## BMI          0.003693   0.003573   1.033    0.301
...
```

> How would I test whether there is a relationship between age and the number of cigarettes smoked, after accounting for other variables?

Inference with negative binomial models

```
...
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.877771   0.123477  23.306  < 2e-16 ***
## male          0.459148   0.027641  16.611  < 2e-16 ***
## age          -0.007010   0.001731  -4.050 5.12e-05 ***
## education2    0.024518   0.032534   0.754    0.451
## education3    0.009252   0.040802   0.227    0.821
## education4   -0.027732   0.044825  -0.619    0.536
## diabetes     -0.010124   0.099126  -0.102    0.919
## BMI           0.003693   0.003573   1.033    0.301
...
```

> How would I test whether there is a relationship between education and the number of cigarettes smoked, after accounting for other variables?

Likelihood ratio test

```
m2 <- glm.nb(cigsPerDay ~ male + age + education +
               diabetes + BMI, data = smokers)
m3 <- glm.nb(cigsPerDay ~ male + age +
               diabetes + BMI, data = smokers)
m2$twologlik - m3$twologlik
```

```
## [1] 1.423055
```

```
pchisq(1.423, df=3, lower.tail=F)
```

```
## [1] 0.7001524
```