Lecture 17

Inference with quasi-Poisson models

An alternative to quasi-Poisson

Poisson:

- Mean = λ_i
- Variance = λ_i

quasi-Poisson:

- Mean = λ_i
- Variance = $\varphi\lambda_i$
- Variance is a linear function of the mean

Question: What if we want variance to depend on the mean in a different way?

The negative binomial distribution

If $Y_i \sim NB(r, p)$, then Y_i takes values $y = 0, 1, 2, 3, \ldots$ with probabilities

$$P(Y_i = y) = \frac{\Gamma(y + r)}{\Gamma(y + 1)\Gamma(r)}(1 - p)^r p^y$$

- $r > 0, \quad p \in [0, 1]$
- $\mathbb{E}[Y_i] = \dfrac{pr}{1 - p} = \mu$
- $Var(Y_i) = \dfrac{pr}{(1 - p)^2} = \mu + \dfrac{\mu^2}{r}$
- Variance is a *quadratic* function of the mean

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

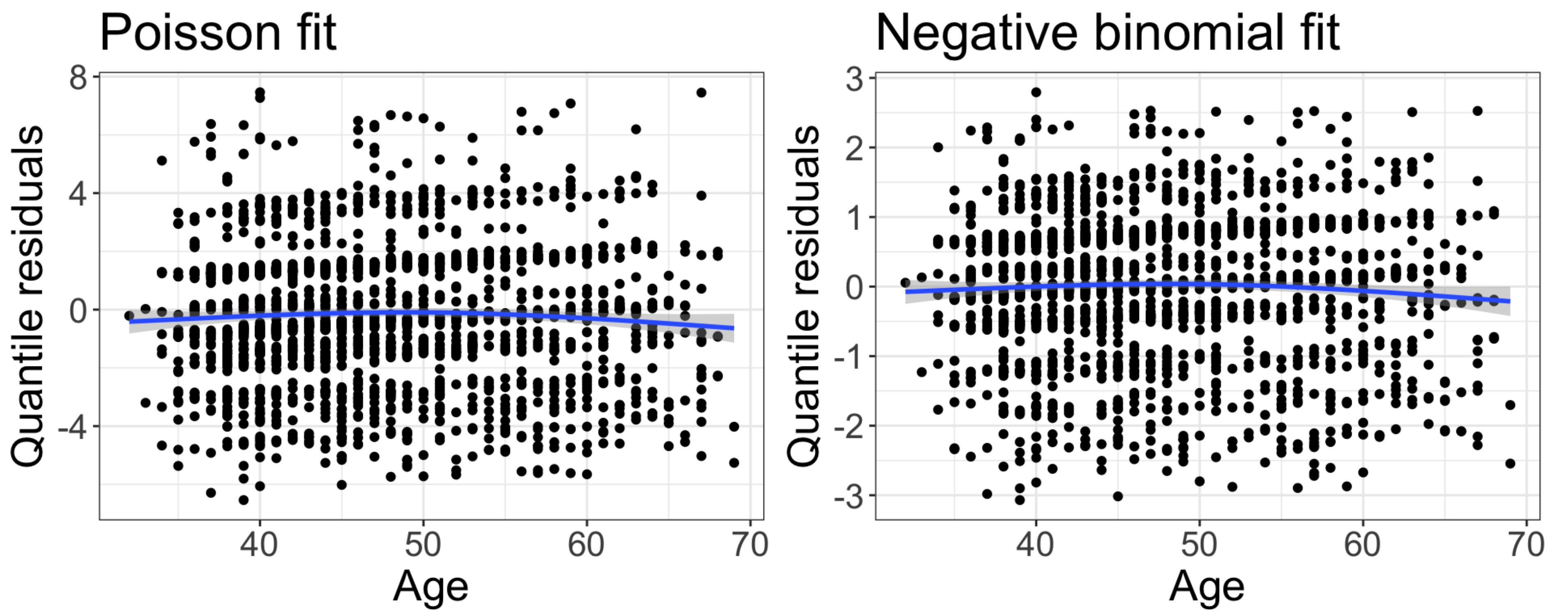
```r
library(MASS)
m2 <- glm.nb(cigsPerDay ~ male + age + education +
                diabetes + BMI, data = smokers)
```

```
...
(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
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(3.2981) family taken to be
1)
```

$$\hat{r} = 3.3$$

Poisson vs. negative binomial fits



Inference with negative binomial models

```
...
(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    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

...
```

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

```
...
(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
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

...
```

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

```r
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
```

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

```
[1] 0.7001524
```

Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_17.html