Lecture 15

Motivating example: air pollution data

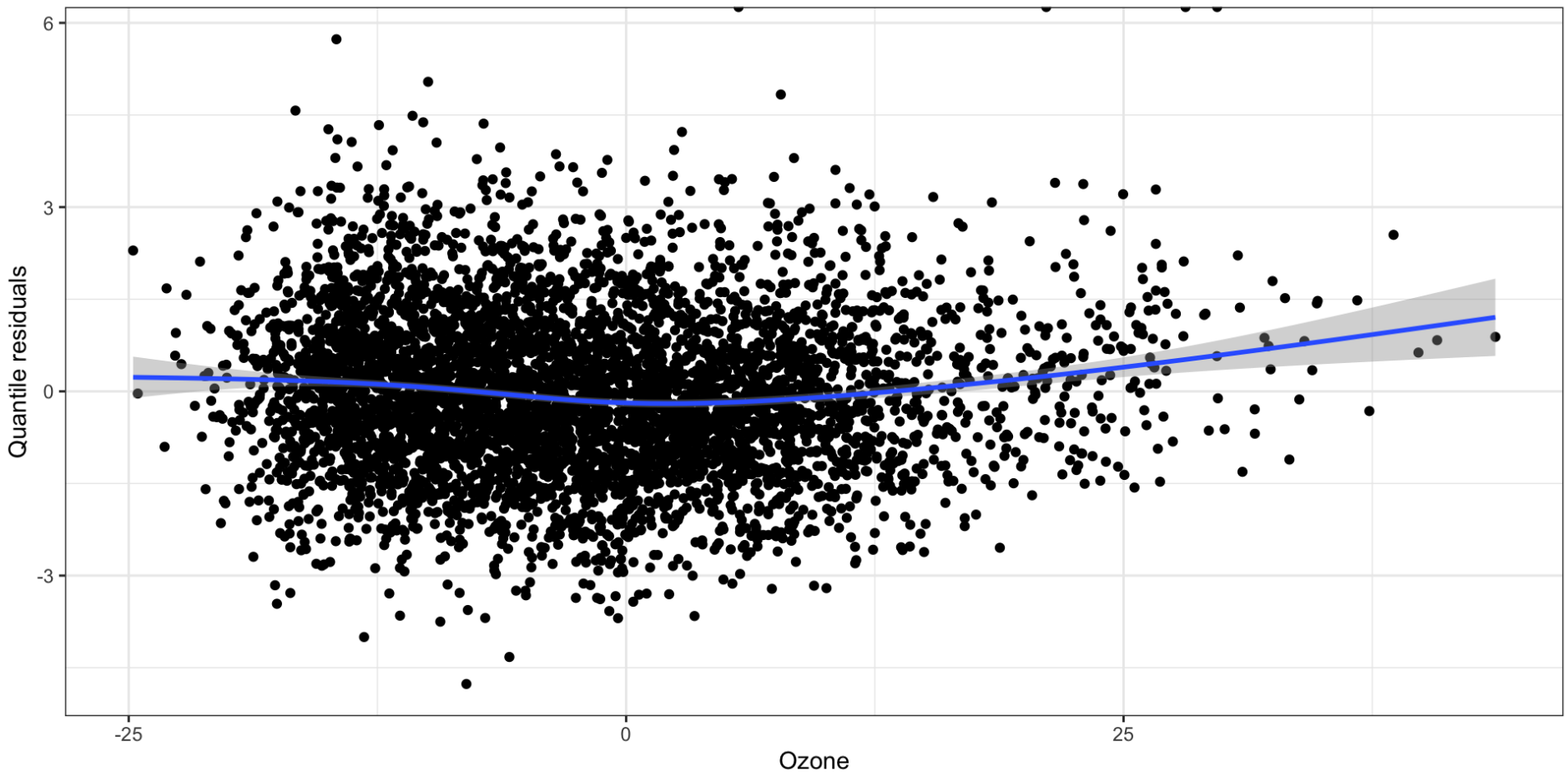
- Data on Chicago air quality and death between 1987 and 2000
- Variables include:
  - deaths
  - ozone concentration
  - sulphur dioxide concentration
  - temperature

Motivating example: air pollution data

$$\text{Deaths}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{Ozone}_i$$

Quantile residual plot



GOF test

```
1 m1$deviance
```

```
[1] 9551.836
```

```
1 m1$df.residual
```

```
[1] 5112
```

```
1 pchisq(m1$deviance, m1$df.residual, lower.tail=F)
```

```
[1] 6.362106e-273
```

Overdisperion

Overdispersion occurs when the response variable Y_i has greater variability than the model accounts for

Recap: sandwich estimator for GLMs

Assumptions about both mean and variance

Quasi-Poisson models

Example: Chicago air quality

Poisson model:

```
                Estimate   Std. Error    z value     Pr(>|z|)
(Intercept)  4.743277988 0.0013382057 3544.50583 0.000000e+00
o3median    -0.002301345 0.0001285909  -17.89664 1.252641e-71
```

Quasi-Poisson model:

```
Call:
glm(formula = death ~ o3median, family = quasipoisson, data = chicago)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.7432780  0.0018822 2520.02   <2e-16 ***
o3median    -0.0023013  0.0001809  -12.72   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.978347)

    Null deviance: 9873.8  on 5113  degrees of freedom
```

Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_15.html