Lecture 20

Data from last time

Survey data from 77 college students on a dry campus (i.e., alcohol is prohibited) in the US. Survey asks students "How many alcoholic drinks did you consume last weekend?"

- `drinks`: number of drinks the student reports consuming
- `sex`: whether the student identifies as male
- `OffCampus`: whether the student lives off campus
- `FirstYear`: whether the student is a first-year student

Our goal: model the number of drinks students report consuming.

Recap: Poisson hurdle model

Fitting Poisson hurdle models

Fitting the model in R

```r
library(pscl)

m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,
             dist = "poisson", zero.dist = "binomial",
             data = wdrinks)

m1$coefficients
```

```
$count
  (Intercept)          sexm FirstYearTRUE OffCampusTRUE
    0.8132113     0.9706640    -0.2181068     0.3762608

$zero
  (Intercept)          sexm FirstYearTRUE OffCampusTRUE
    0.1230510     0.3377969    -0.8554289     1.5803472
```

Model assumptions

```r
m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,
             dist = "poisson", zero.dist = "binomial",
             data = wdrinks)

m1$coefficients
```

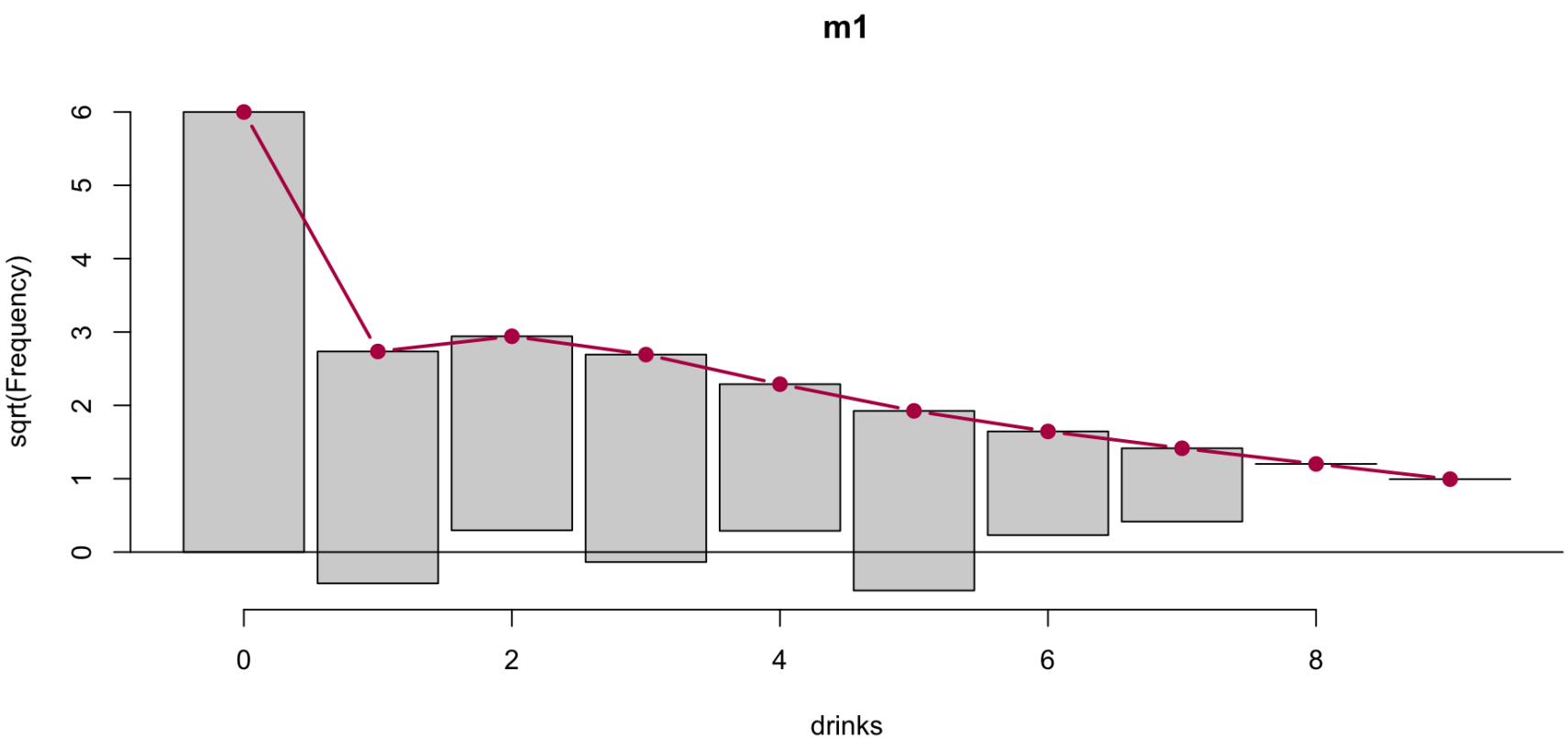
```
$count
  (Intercept)          sexm FirstYearTRUE OffCampusTRUE
    0.8132113     0.9706640    -0.2181068     0.3762608

$zero
  (Intercept)          sexm FirstYearTRUE OffCampusTRUE
    0.1230510     0.3377969    -0.8554289     1.5803472
```

What assumptions does this model make?

Model diagnostics

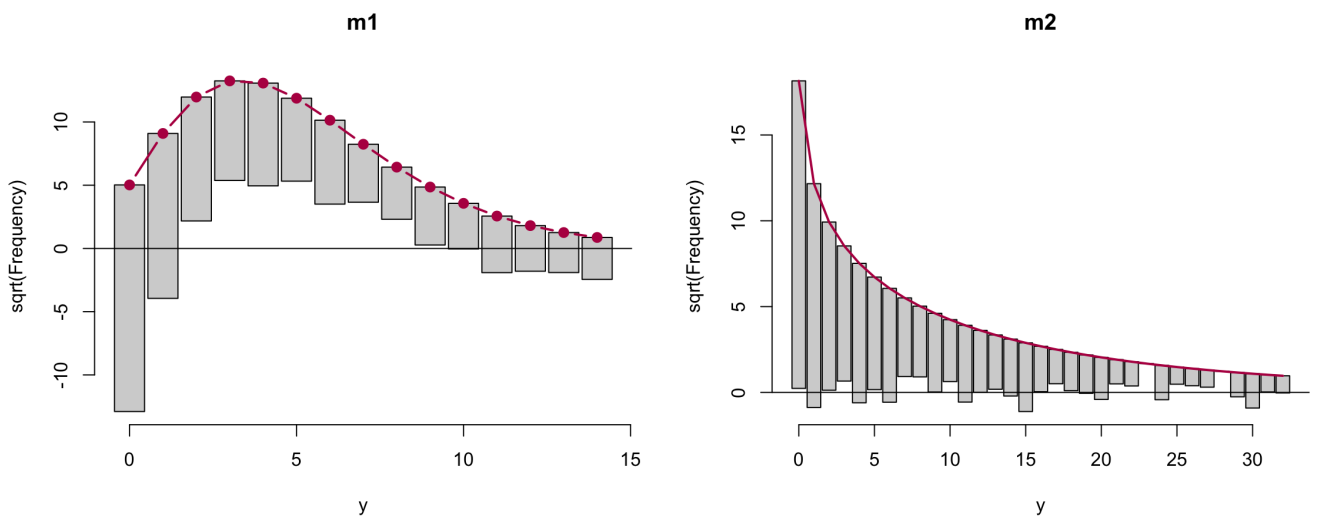
Rootograms

```
1 library(countreg)
2 rootogram(m1)
```



Other examples with rootograms

```
par(mfrow=c(1, 2))

x <- rnorm(1000)
y <- rnbinom(1000, 0.5, mu=exp(1.5 + 0.2*x))
m1 <- glm(y ~ x, family = poisson)
m2 <- glm.nb(y ~ x)

rootogram(m1)
rootogram(m2)
```



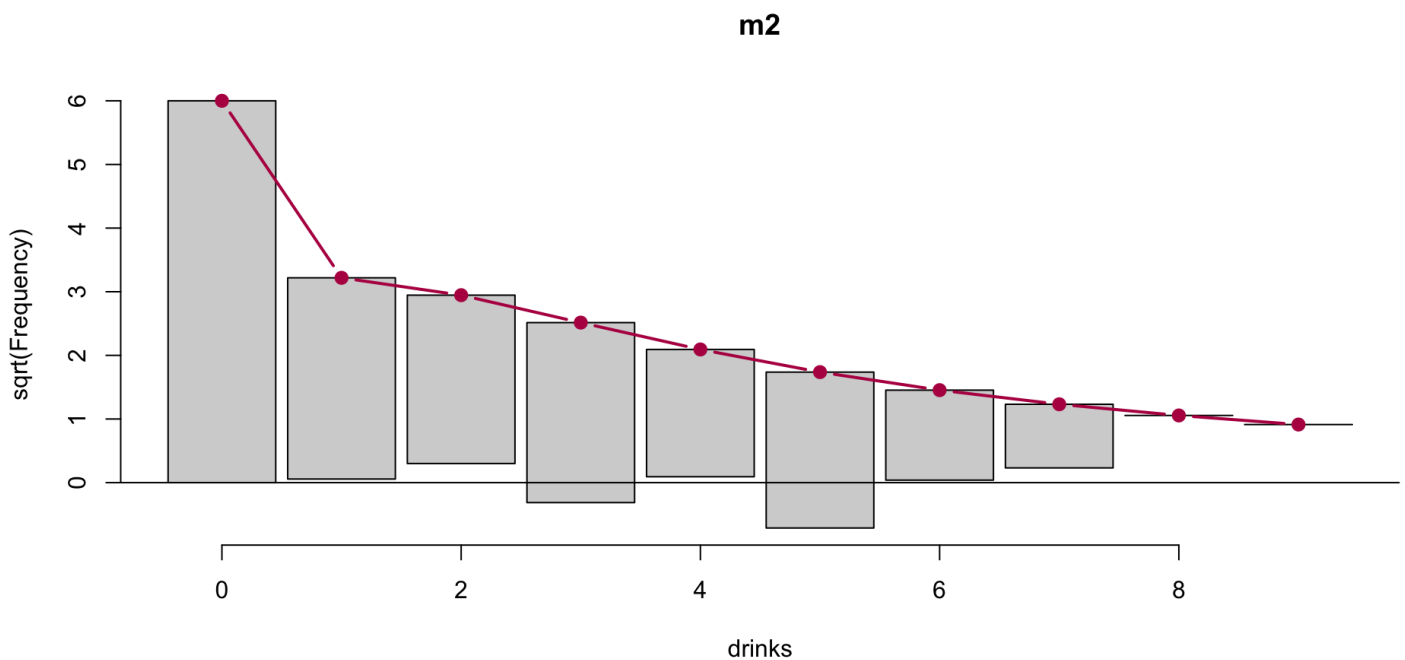
Hurdle models for count data

$$P(Y_i > 0) = p_i \quad g_{zero}(p_i) = \gamma^T X_i$$

$$Y_i | (Y_i > 0) \sim ZT(\lambda_i) \quad g_{count}(\lambda_i) = \beta^T X_i$$

Negative binomial hurdle model

```r
m2 <- hurdle(drinks ~ sex + FirstYear + OffCampus,
             dist = "negbin", zero.dist = "binomial",
             data = wdrinks)

rootogram(m2)
```



Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_20.html