Lecture 19

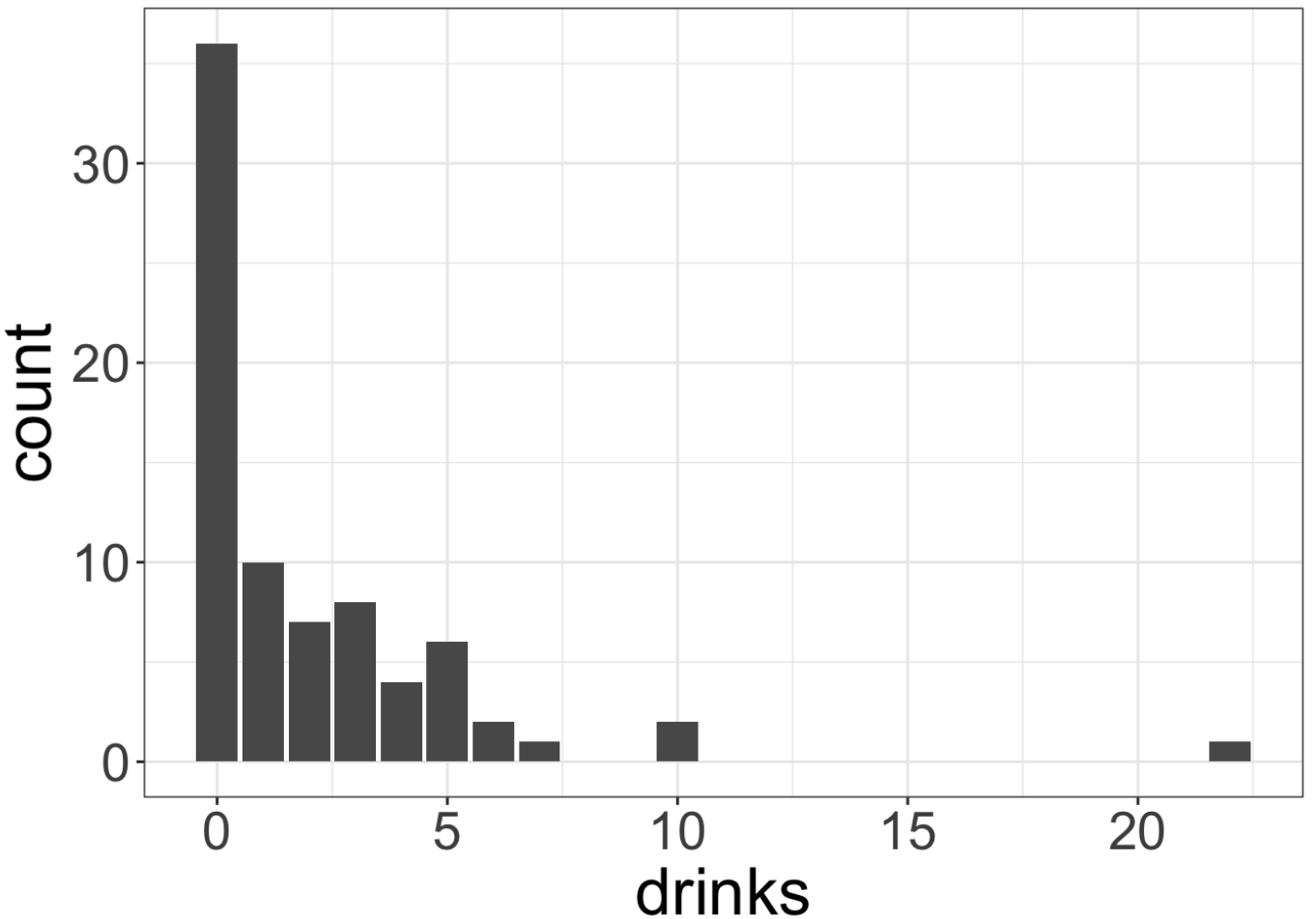
New data

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

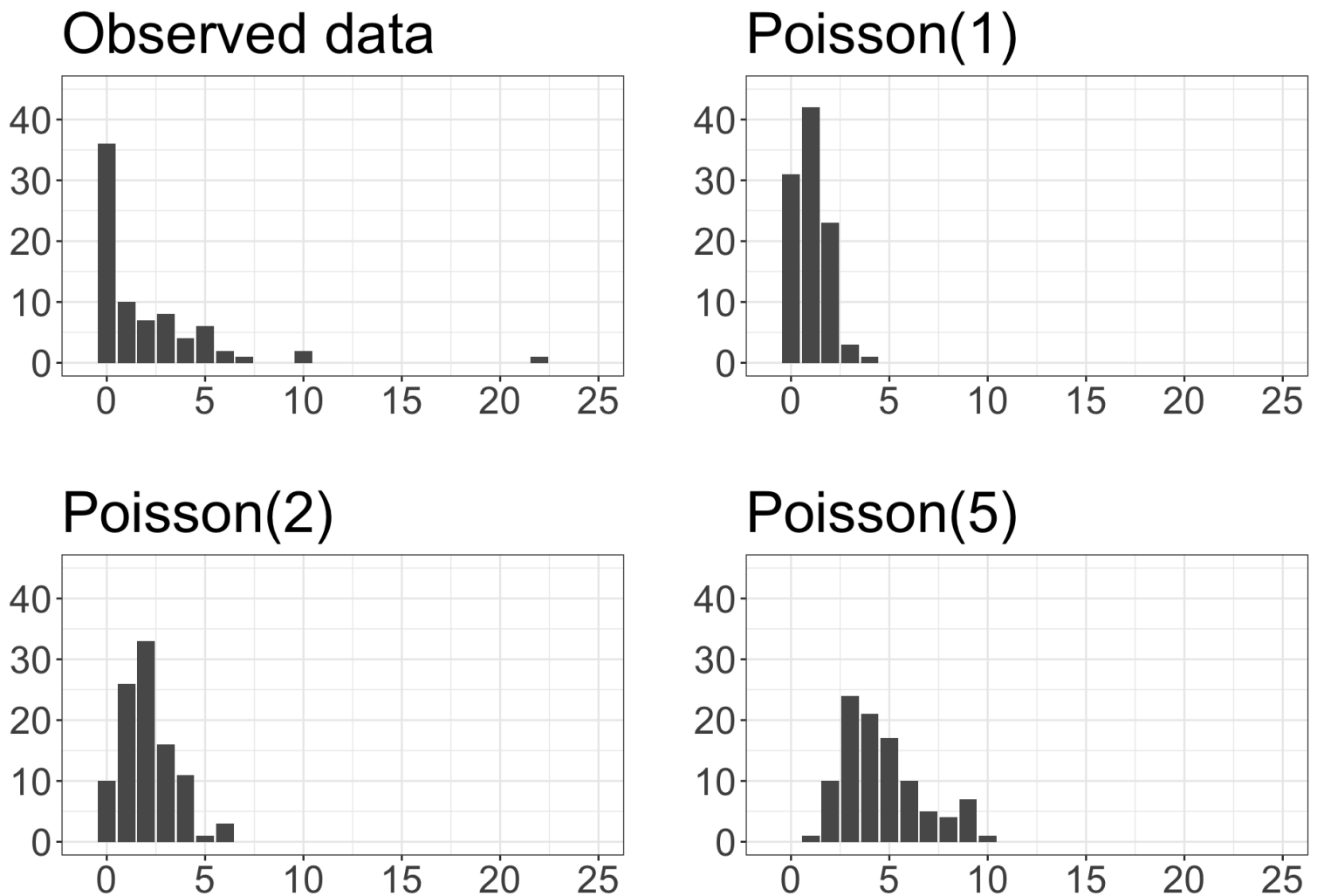
Our goal: model the number of drinks students report consuming.

EDA: drinks



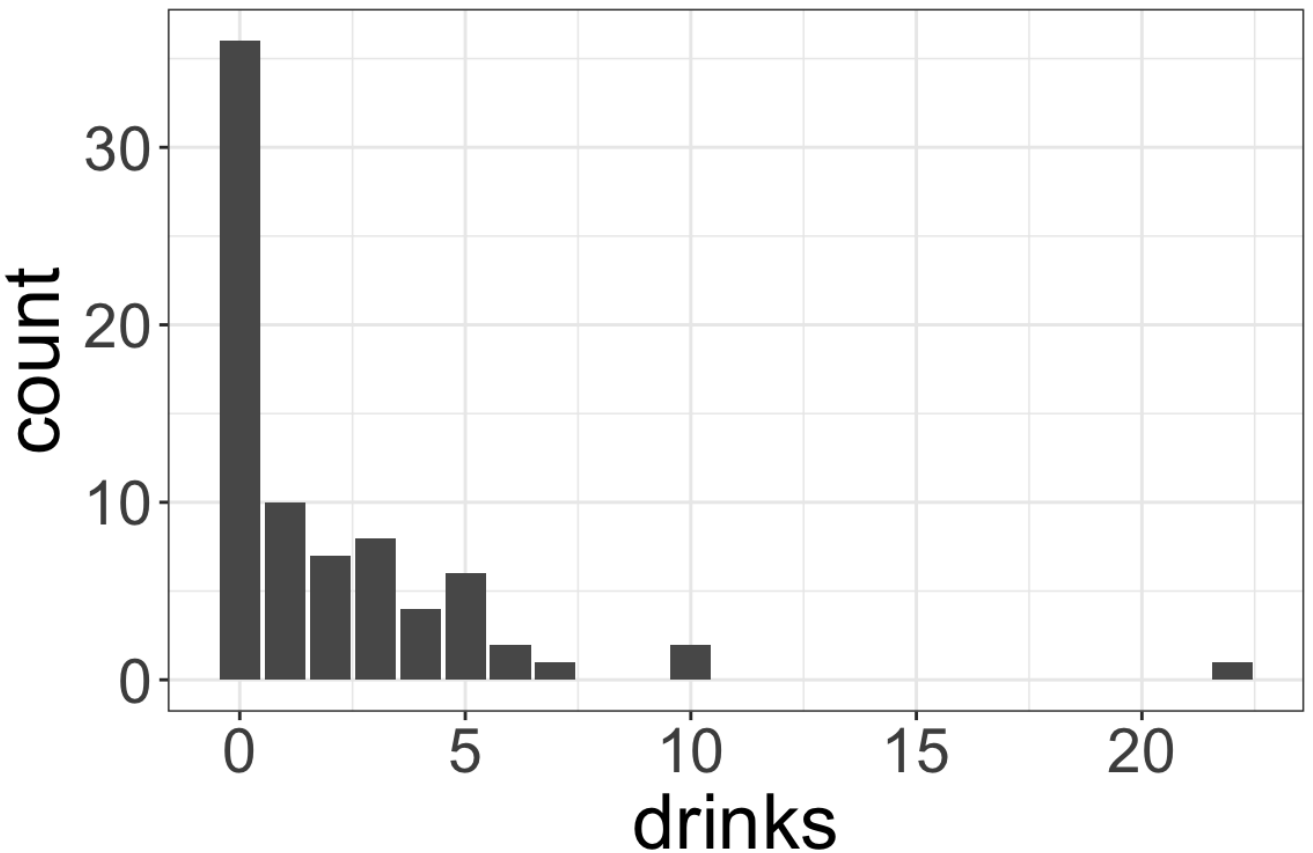
What do you notice about this distribution?

Comparisons with Poisson distributions



Excess zeros

Why might there be excess 0s in the data, and why is that a problem for modeling the number of drinks consumed?



Hurdle models: model the zeros separately

Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_19.html