Welcome!

https://github.com/lauken13/Beginners_Bayes_Workshop

Outline

Tuesday AM

- Introduce our running example
- Bayesian workflow
- Code a simple model with Stan

Tuesday PM

- Code a more complex model with Stan
- Run the workflow

Wednesday AM

- Hierarchical modeling with Stan
- How does HMC-NUTS work?

Wednesday PM

- Hierarchical modeling with Stan (part 2)
- How to think about warnings/errors?
- More resources

Why do Bayesian Inference?

- More complex models that reflect the believed generative process
- Regularization with priors
- Easier quantification of uncertainty

Why use Stan?

- Sampling algorithm
- Diagnostics
- Community
- Resources
- Quality

Data

- 10 buildings in NYC (b)
- Cockroach traps are randomly distributed amongst buildings every month
- At the end of the month, number of complaints are tallied (t)
- Some seasonality (available in case study)
- Overall Q: Optimal number of traps (available in case study)
- This conference: What is the relationship between the number of traps and the number of complaints

Data

- Additional information about the building
  - Number of floors
  - Square footage of building
  - Log square footage of building
  - Live in super
  - Average rent
  - Average tenant age
  - Building age

Simple model

$$\textbf{complaints}_{b,t} \sim \textbf{Poisson}(\lambda_{b,t})$$
$$\lambda_{b,t} = \exp(\eta_{b,t})$$
$$\eta_{b,t} = \alpha + \beta\, \textbf{traps}_{b,t}$$

Bayesian workflow

1. Prior predictive checks
2. Exploratory data analysis
3. Sampling
4. Posterior predictive checks
5. Model comparison
6. Decision analysis

Bayesian workflow

1. **Prior predictive checks**
2. **Exploratory data analysis**
3. Sampling (day 2!)
4. **Posterior predictive checks**
5. Model comparison (see Aki's workshop & vignettes)
6. Decision analysis (see vignette)

Prior Predictive Checks

- Used to understand the role of the prior in the data generating process
- Even people who don't have strong feelings about a specific prior know what their observed data shouldn't look like (i.e., pollution levels aren't so thick that we can't move, so the model should reflect that)

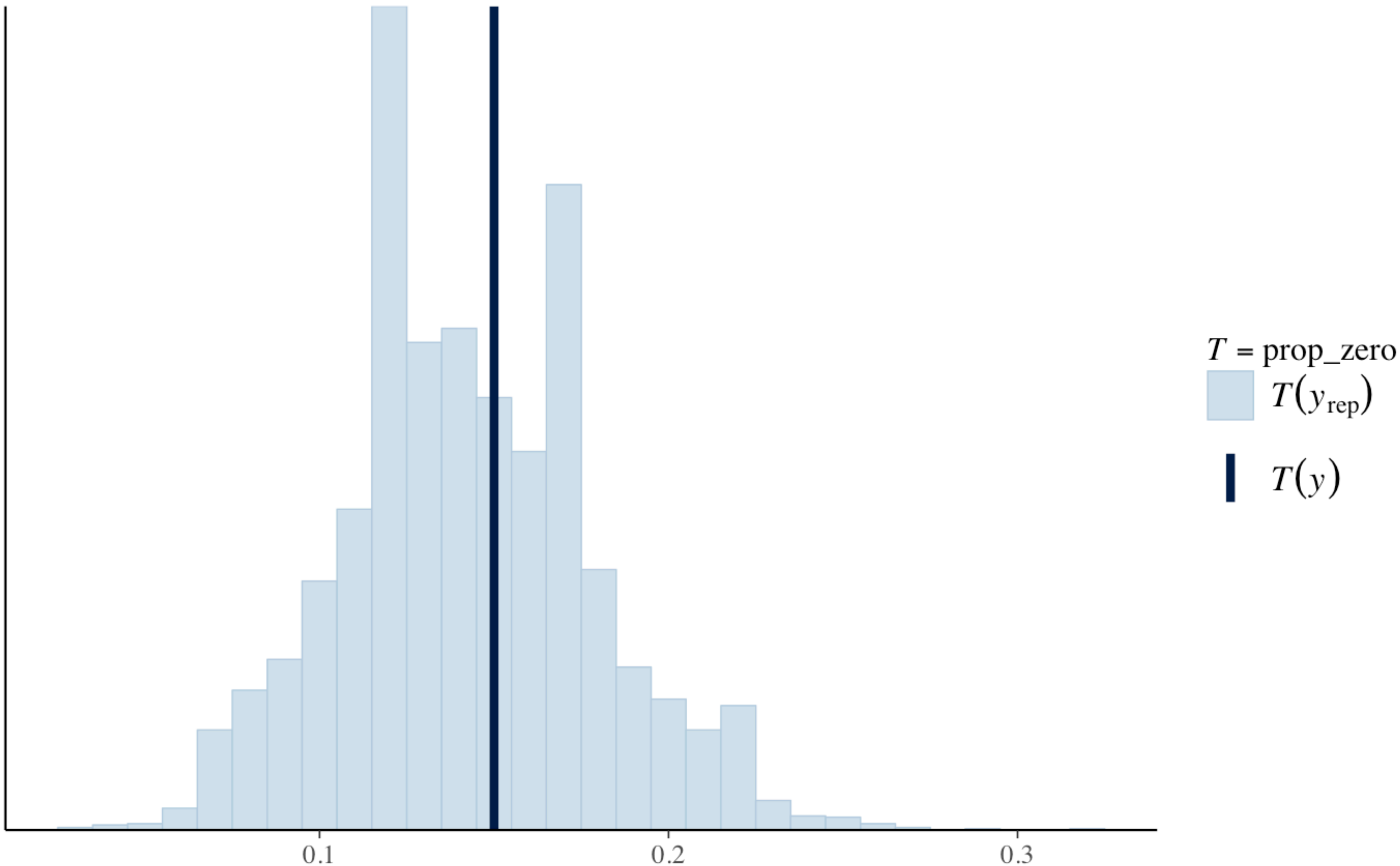
1. Use either observed x or generate reasonable values for x
2. Sample from your prior
3. Calculate the predicted y for that x and that particular random draw
4. Repeat 2 & 3 many times
5. Compute summary statistics if necessary

Coding with Stan

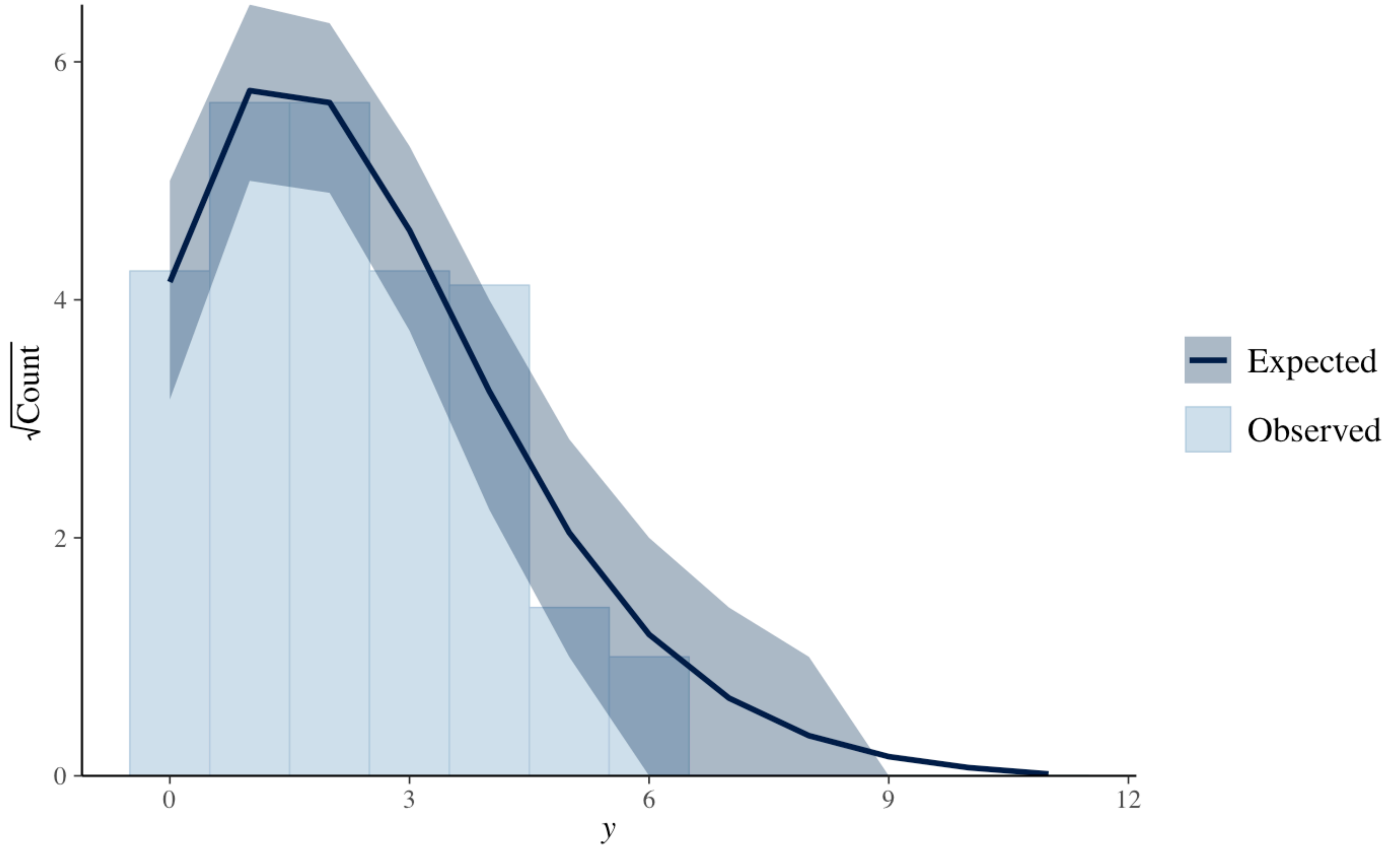
Posterior Predictive Check

- Used to understand the fit of the model
- The predicted distribution
  1. Sample from your posterior
  2. Use observed x
  3. Calculate the predicted y for that x
  4. Repeat many times
  5. Compute summary statistics if necessary

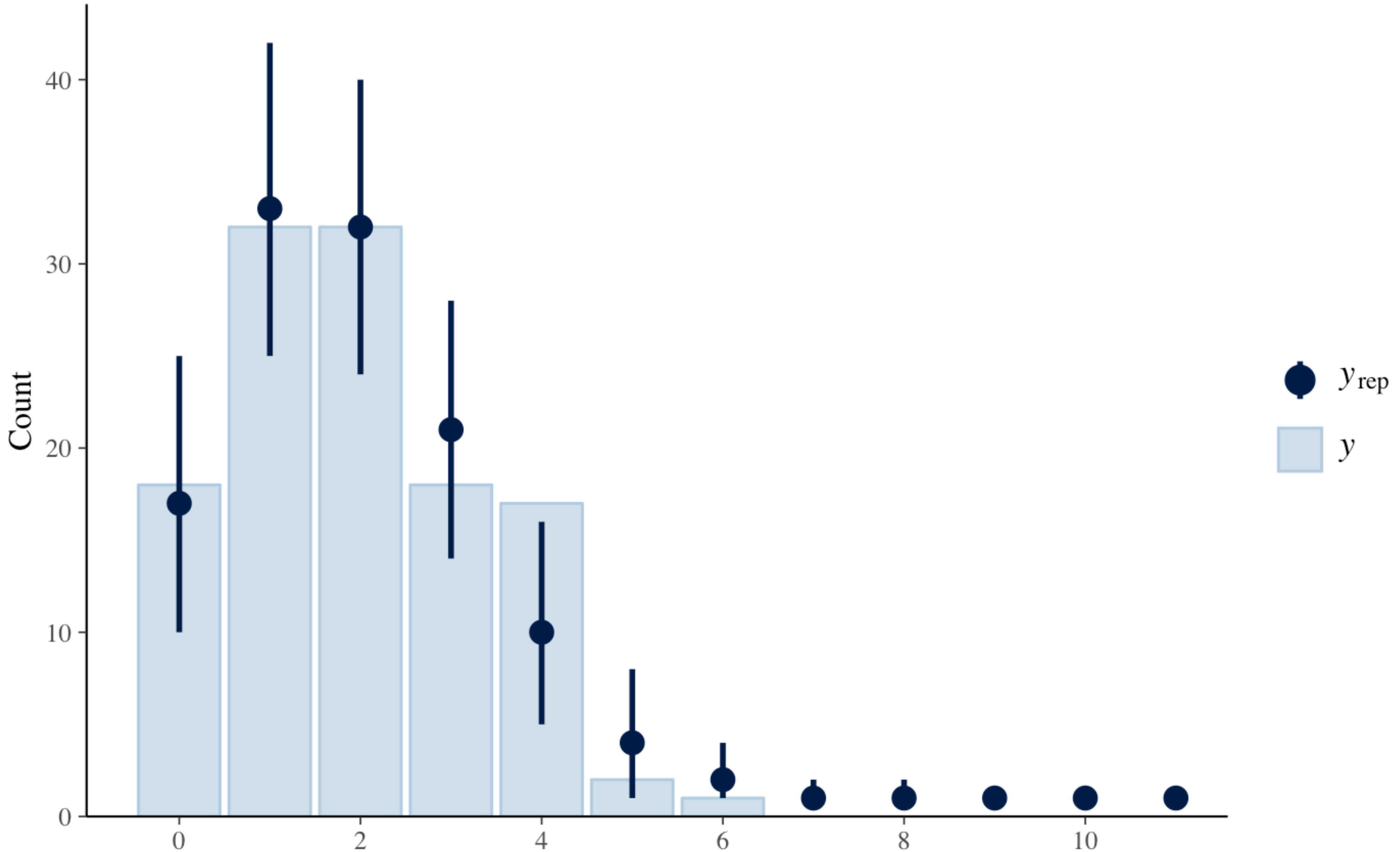
Posterior Predictive Check



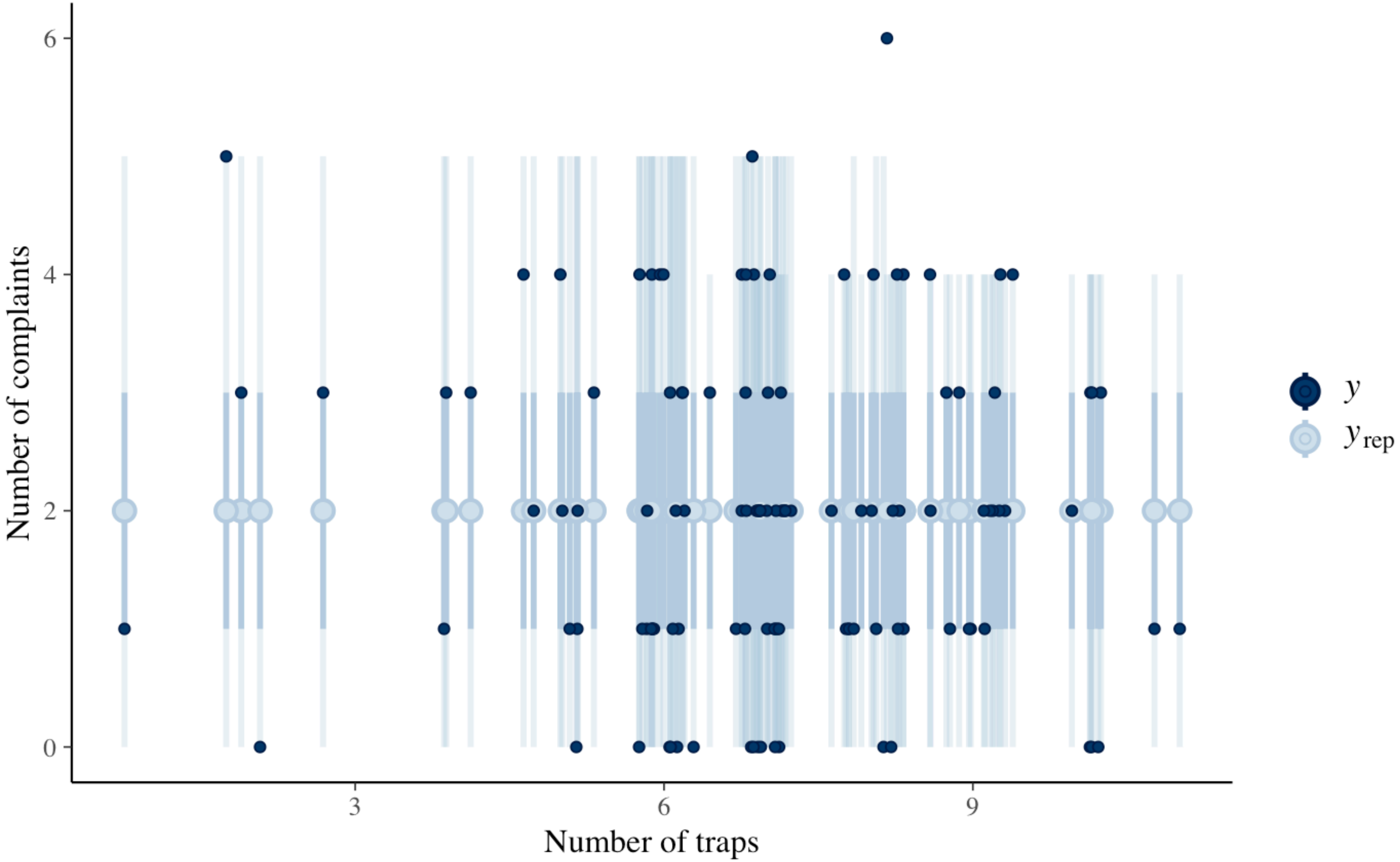
Posterior Predictive Check



Posterior Predictive Check



Posterior Predictive Check



Hierachical models

$$\begin{aligned}
\text{complaints}_{b,t} &\sim \text{Neg-Binomial}(\lambda_{b,t}, \phi) \\
\lambda_{b,t} &= \exp\left(\eta_{b,t}\right) \\
\eta_{b,t} &= \alpha + \beta \, \text{traps}_{b,t} + \beta_{\text{super}} \, \text{super}_b + \text{log_sq_foot}_b
\end{aligned}$$

Hierachical models

$$
\begin{aligned}
\text{complaints}_{b,t} &\sim \text{Neg-Binomial}(\lambda_{b,t}, \phi) \\
\lambda_{b,t} &= \exp\left(\eta_{b,t}\right) \\
\eta_{b,t} &= \alpha + \beta\,\text{traps}_{b,t} + \beta_{\text{super}}\,\text{super}_b + \text{log\_sq\_foot}_b
\end{aligned}
$$

$$
\begin{gathered}
\text{complaints}_{b,t} \sim \text{Neg-Binomial}(\lambda_{b,t}, \phi) \\
\lambda_{b,t} = \exp\left(\eta_{b,t}\right) \\
\eta_{b,t} = \mu_b + \beta\,\text{traps}_{b,t} + \beta_{\text{super}}\,\text{super}_b + \text{log\_sq\_foot}_b \\
\mu_b \sim \text{Normal}(\alpha, \sigma_\mu)
\end{gathered}
$$

Hierachical models

$$\text{complaints}_{b,t} \sim \text{Neg-Binomial}(\lambda_{b,t}, \phi)$$
$$\lambda_{b,t} = \exp\left(\eta_{b,t}\right)$$
$$\eta_{b,t} = \mu_b + \beta\,\text{traps}_{b,t} + \beta_{\text{super}}\,\text{super}_b + \text{log\_sq\_foot}_b$$
$$\mu_b \sim \text{Normal}(\alpha, \sigma_\mu)$$

Hierachical models

$$\text{complaints}_{b,t} \sim \text{Neg-Binomial}(\lambda_{b,t}, \phi)$$
$$\lambda_{b,t} = \exp\left(\eta_{b,t}\right)$$
$$\eta_{b,t} = \mu_b + \beta\,\text{traps}_{b,t} + \beta_{\text{super}}\,\text{super}_b + \text{log\_sq\_foot}_b$$
$$\mu_b \sim \text{Normal}(\alpha, \sigma_\mu)$$

$$\eta_{b,t} = \mu_b + \beta\,\text{traps}_{b,t} + \text{log\_sq\_foot}_b$$
$$\mu_b \sim \text{Normal}(\alpha + \beta_{\text{super}}\,\text{super}_b, \sigma_\mu)$$

Hierachical models

$$\text{complaints}_{b,t} \sim \text{Neg-Binomial}(\lambda_{b,t}, \phi)$$
$$\lambda_{b,t} = \exp\left(\eta_{b,t}\right)$$
$$\eta_{b,t} = \mu_b + \beta \,\text{traps}_{b,t} + \beta_{\text{super}}\,\text{super}_b + \text{log\_sq\_foot}_b$$
$$\mu_b \sim \text{Normal}(\alpha, \sigma_\mu)$$

$$\eta_{b,t} = \mu_b + \beta \,\text{traps}_{b,t} + \text{log\_sq\_foot}_b$$
$$\mu_b \sim \text{Normal}(\alpha + \beta_{\text{super}}\,\text{super}_b, \sigma_\mu)$$

$$\eta_{b,t} = \mu_b + \beta \,\text{traps} + \text{log\_sq\_foot}$$
$$\mu \sim \text{Normal}(\alpha + \texttt{building\_data}\,\zeta, \,\sigma_\mu)$$

Hierachical models

$$\text{complaints}_{b,t} \sim \text{Neg-Binomial}(\lambda_{b,t}, \phi)$$
$$\lambda_{b,t} = \exp\left(\eta_{b,t}\right)$$
$$\eta_{b,t} = \mu_b + \beta\,\text{traps}_{b,t} + \beta_{\text{super}}\,\text{super}_b + \text{log\_sq\_foot}_b$$
$$\mu_b \sim \text{Normal}(\alpha, \sigma_\mu)$$

$$\eta_{b,t} = \mu_b + \beta\,\text{traps} + \text{log\_sq\_foot}$$
$$\mu \sim \text{Normal}(\alpha + \texttt{building\_data}\,\zeta,\ \sigma_\mu)$$