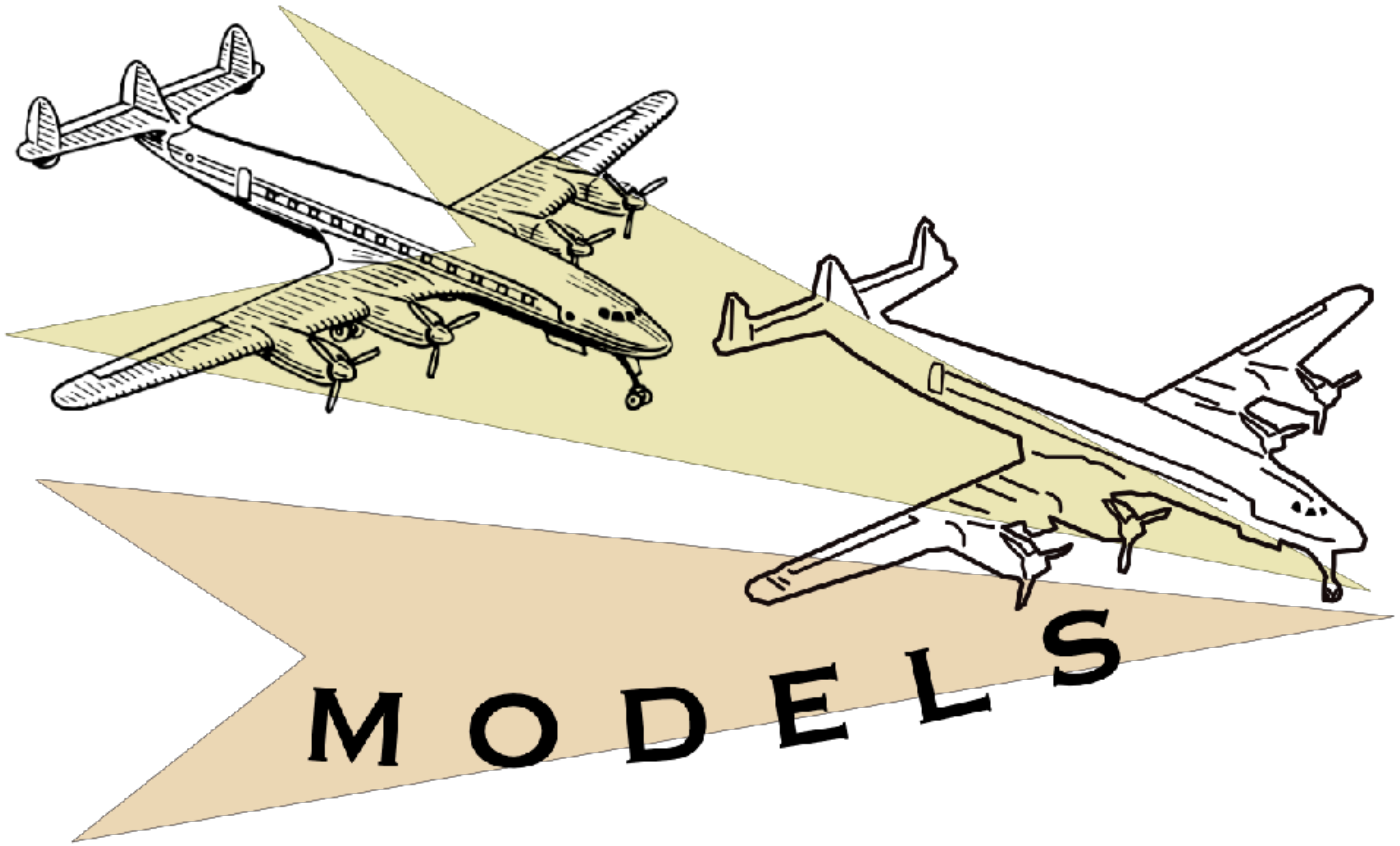
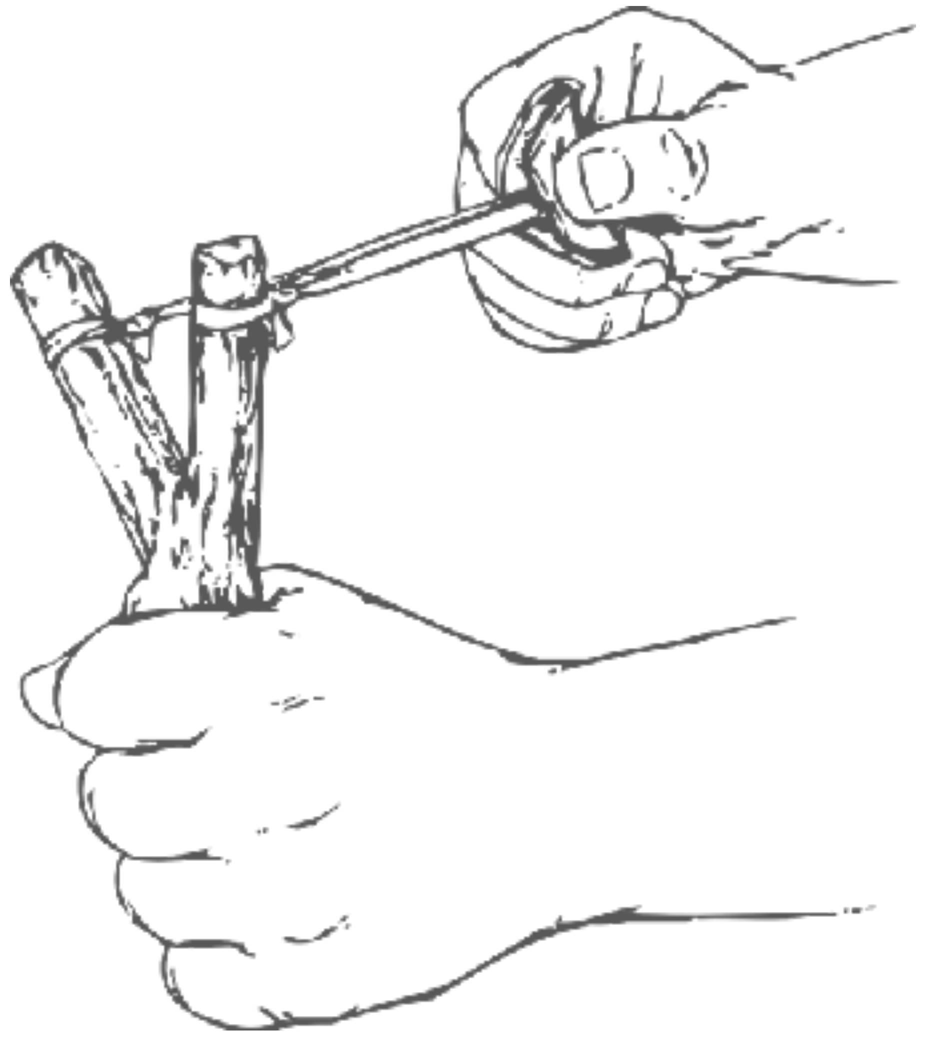
INTRODUCTION TO DATA ANALYSIS

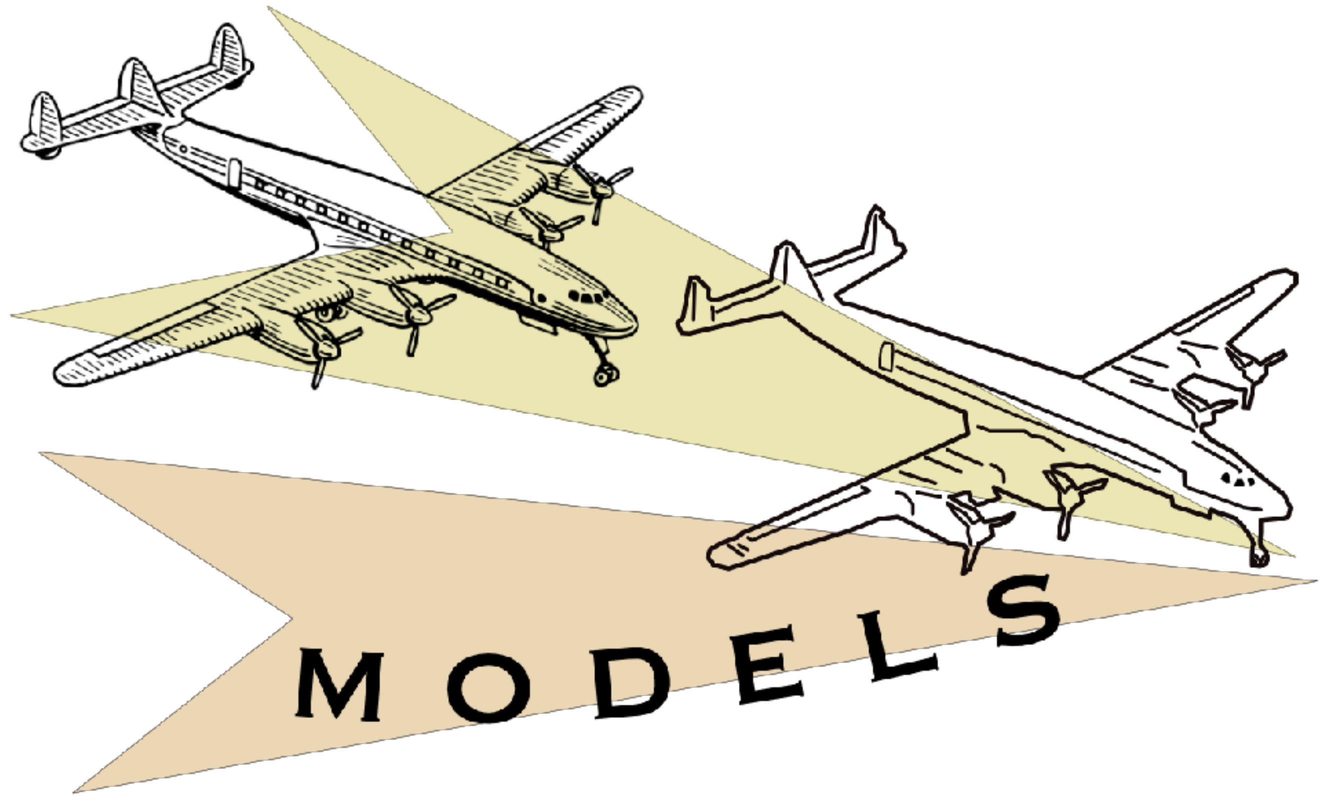
MODELS



LEARNING GOALS



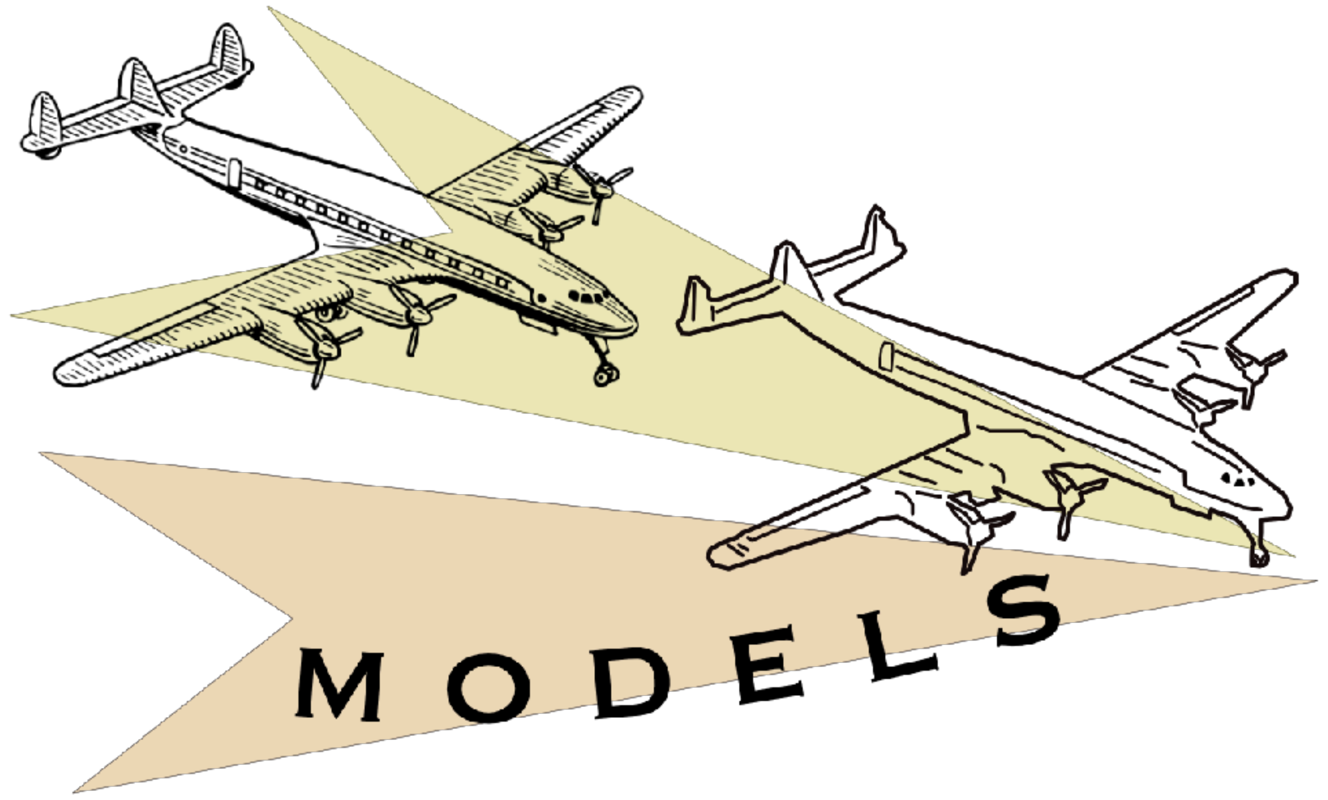
- become acquainted with statistical models
- understand what parameters are and what priors can do
- meet pivotal exemplars:
  - Binomial Model, T-Test Model, Simple Linear Regression
- understand notation to communicate models
  - formulas & graphs



Statistical models

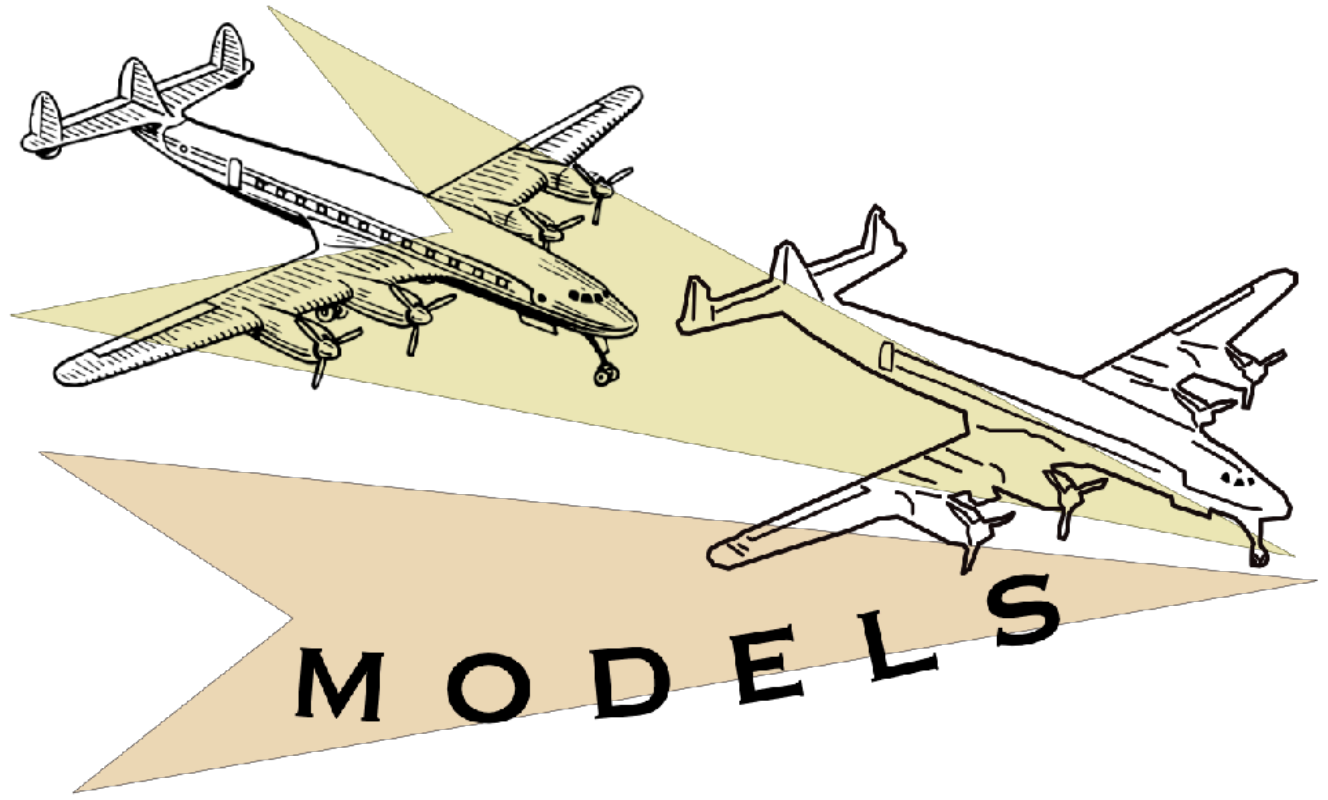
STATISTICAL MODELS

- A statistical model is a condensed formal representation, following common conventional practices of formalization, of the assumptions we make about what the data is and how it might have been generated.



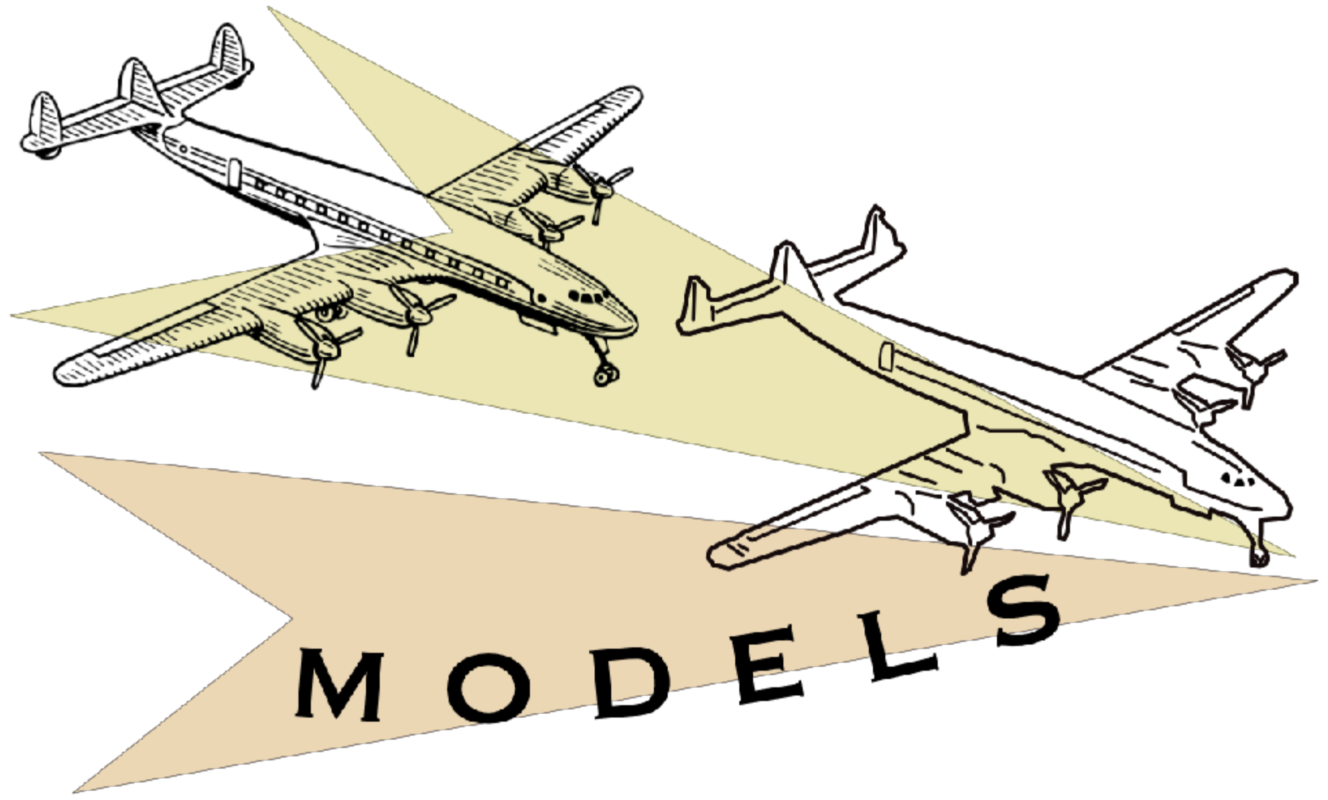
PRAGMATISM IN MODELING

> All models are wrong, but some are useful. — Box (1979)



DEFINITION

- a statistical model M of random process R generating data D consists of:
  - a partition into D_{IV} and D_{DV} of a subset of D
  - a likelihood function: $P_M(D \mid \theta)$
  - [if Bayesian] a prior: $P_M(\theta)$



First examples

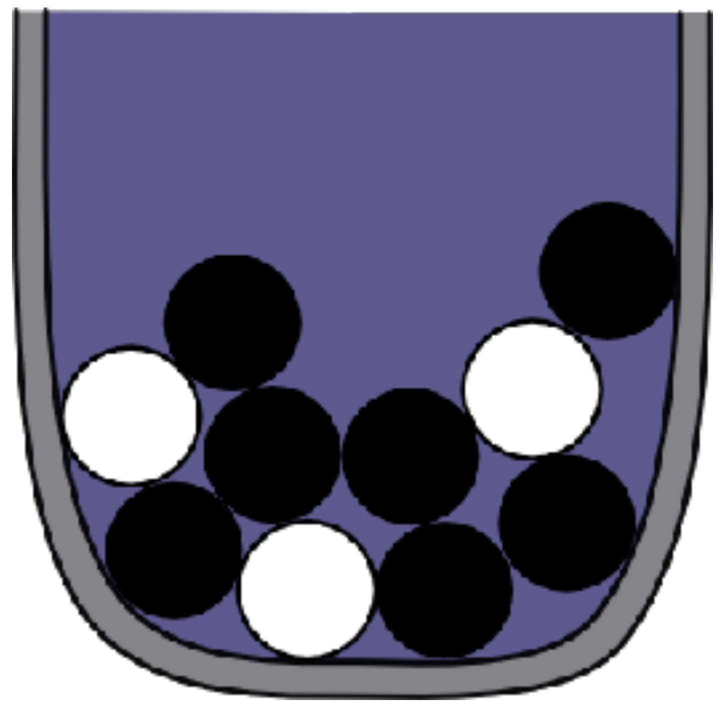
SINGLE DRAW FROM AN URN

- urn contains $N = 10$ balls
- unknown number $k \in \{0,1,\ldots,10\}$ of black balls (rest white)
- likelihood function is uncontroversial

$$P_M(D = \text{black} \mid k) = \frac{k}{N}$$

- possible prior:

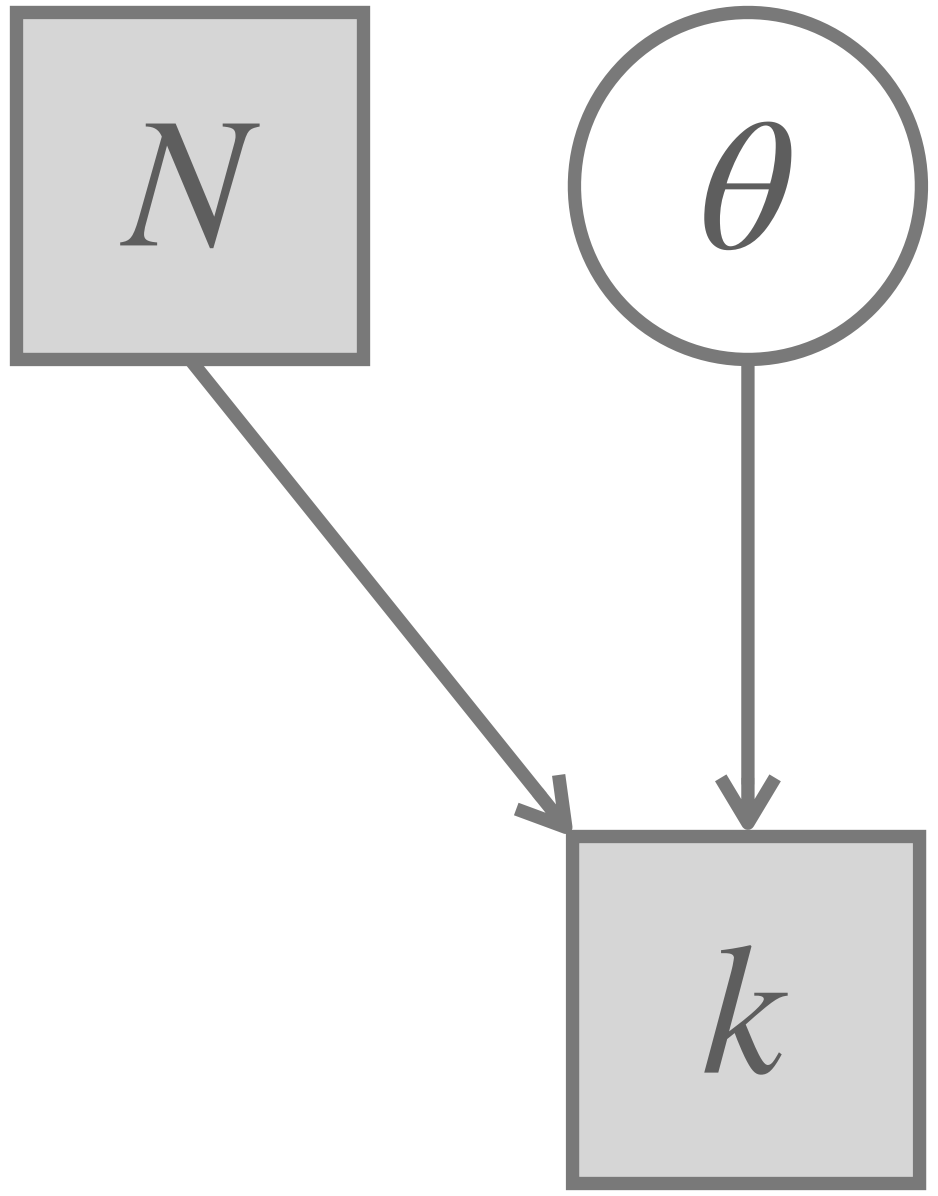
$$P_M(k = i) = \frac{1}{11}, \quad \text{for all } i \in \{0,1,\ldots,10\}$$



one out of eleven possible urns

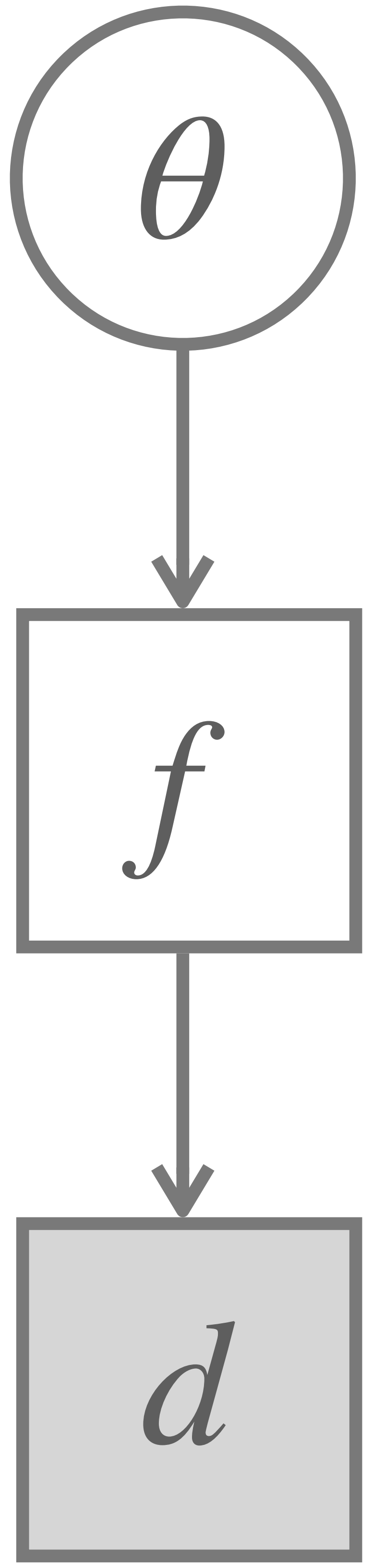
Example models

BINOMIAL MODEL



$$\theta \sim \text{Beta}(\dots)$$
$$k \sim \text{Binomial}(\theta, N)$$

FLIP-AND-DRAW

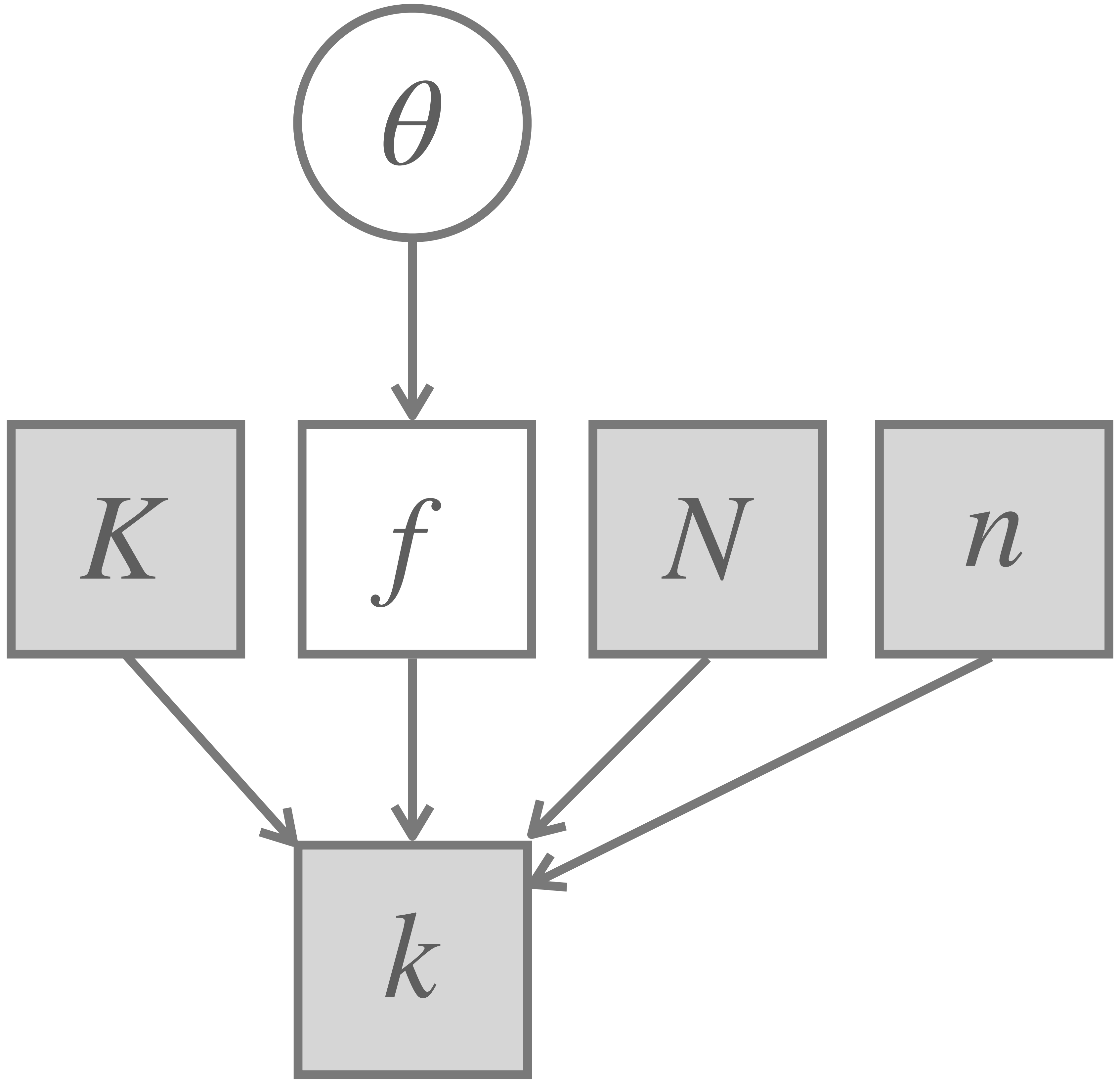


$$\theta \sim \text{Beta}(\ldots)$$

$$f \sim \text{Bernoulli}(\theta)$$

$$d \sim \text{Categorical}(\overrightarrow{p_f})$$

FLIP-AND-DRAW-HYPERGEOMETRIC

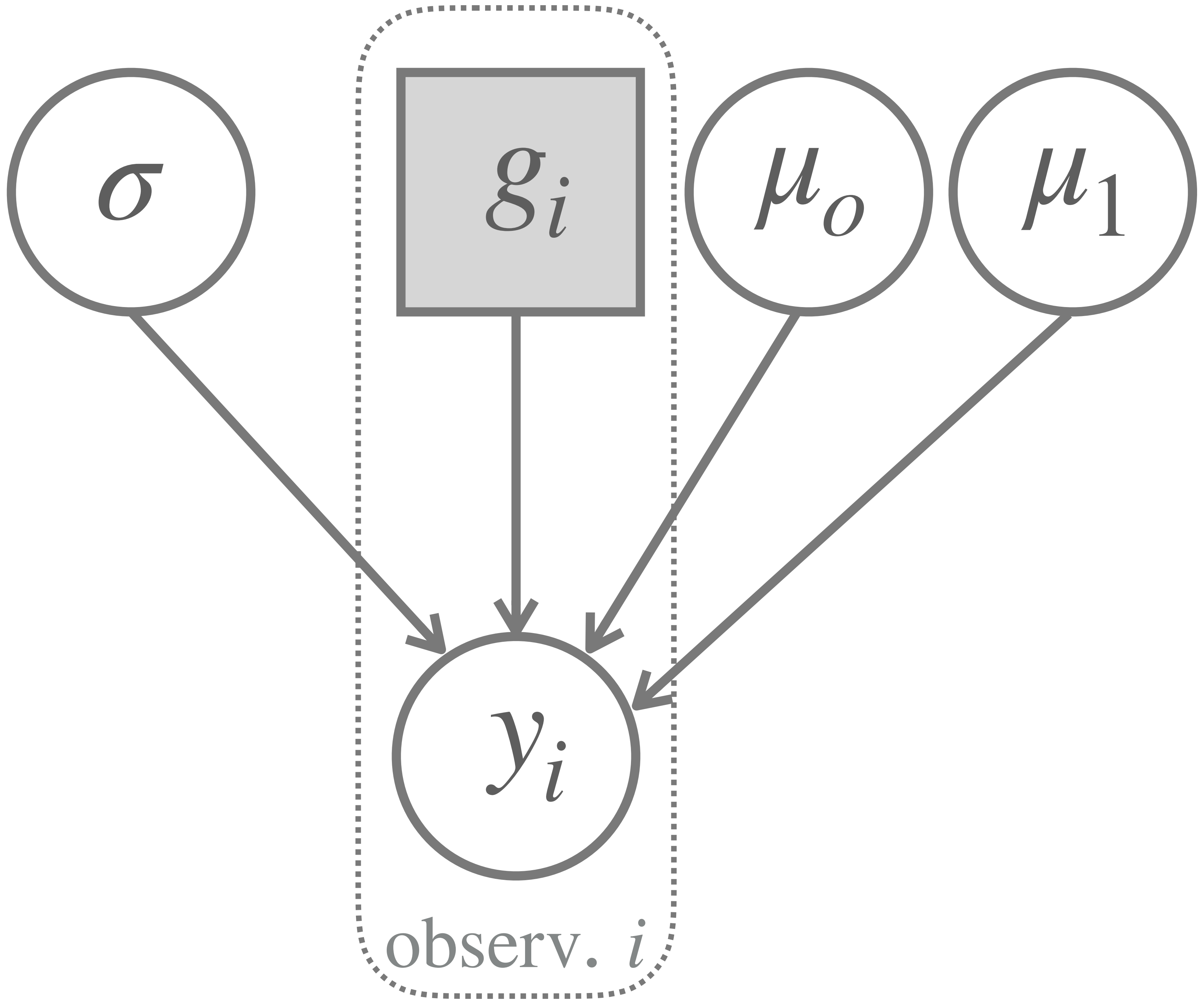


$$\theta \sim \text{Beta}(\ldots)$$

$$f \sim \text{Bernoulli}(\theta)$$

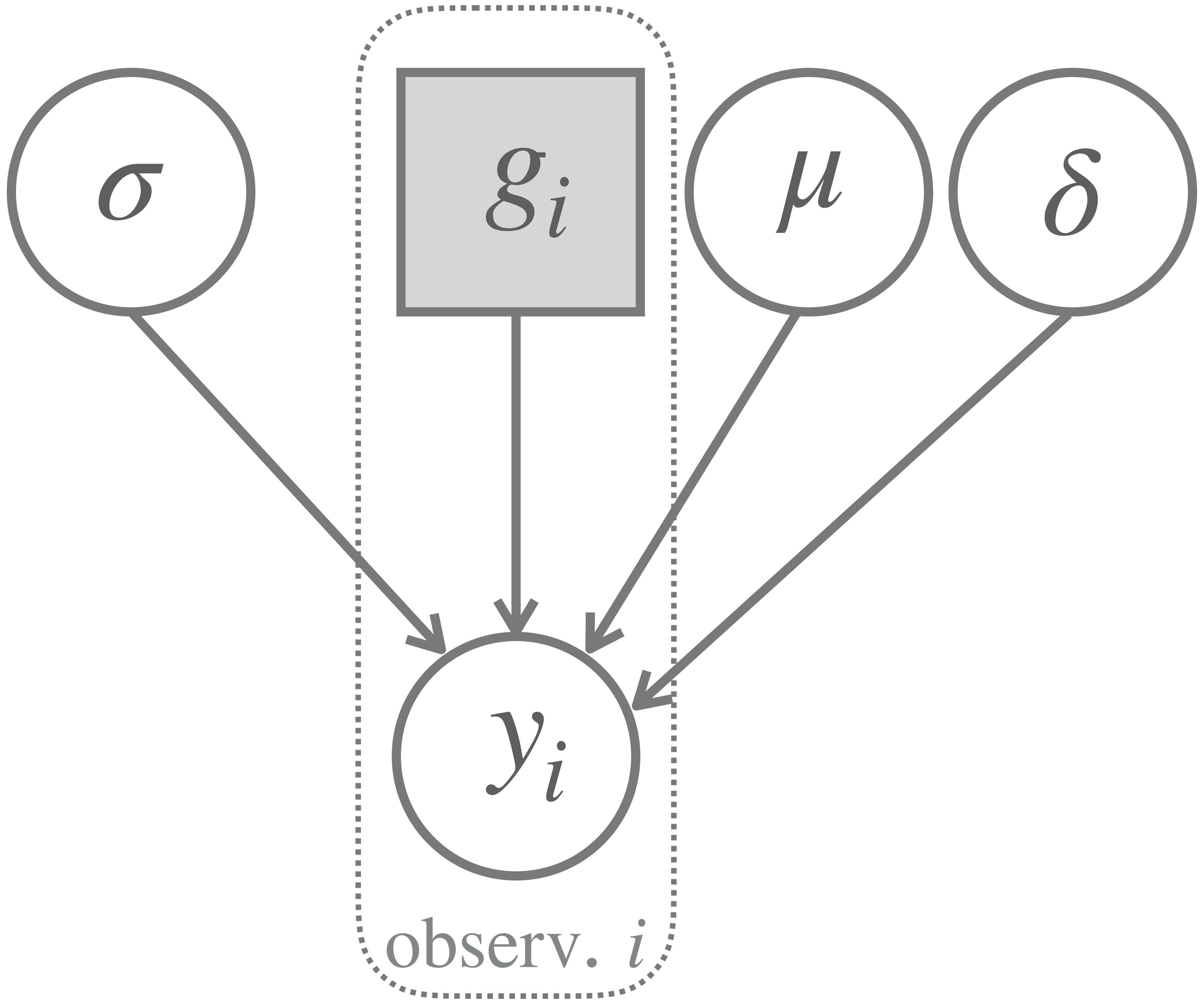
$$k \sim \text{Hypergeometric}(n, N, K_f)$$

T-TEST MODEL [TWO UNCOUPLED MEANS]



$$\sigma \sim \text{Trunc-Norm}(\ldots, \text{lower} = 0)$$
$$\mu_0 \sim \text{Normal}(\ldots)$$
$$\mu_1 \sim \text{Normal}(\ldots)$$
$$y_i \sim \text{Normal}(\mu_{g_i}, \sigma)$$

T-TEST MODEL [WITH DIFFERENCE BETWEEN MEANS]



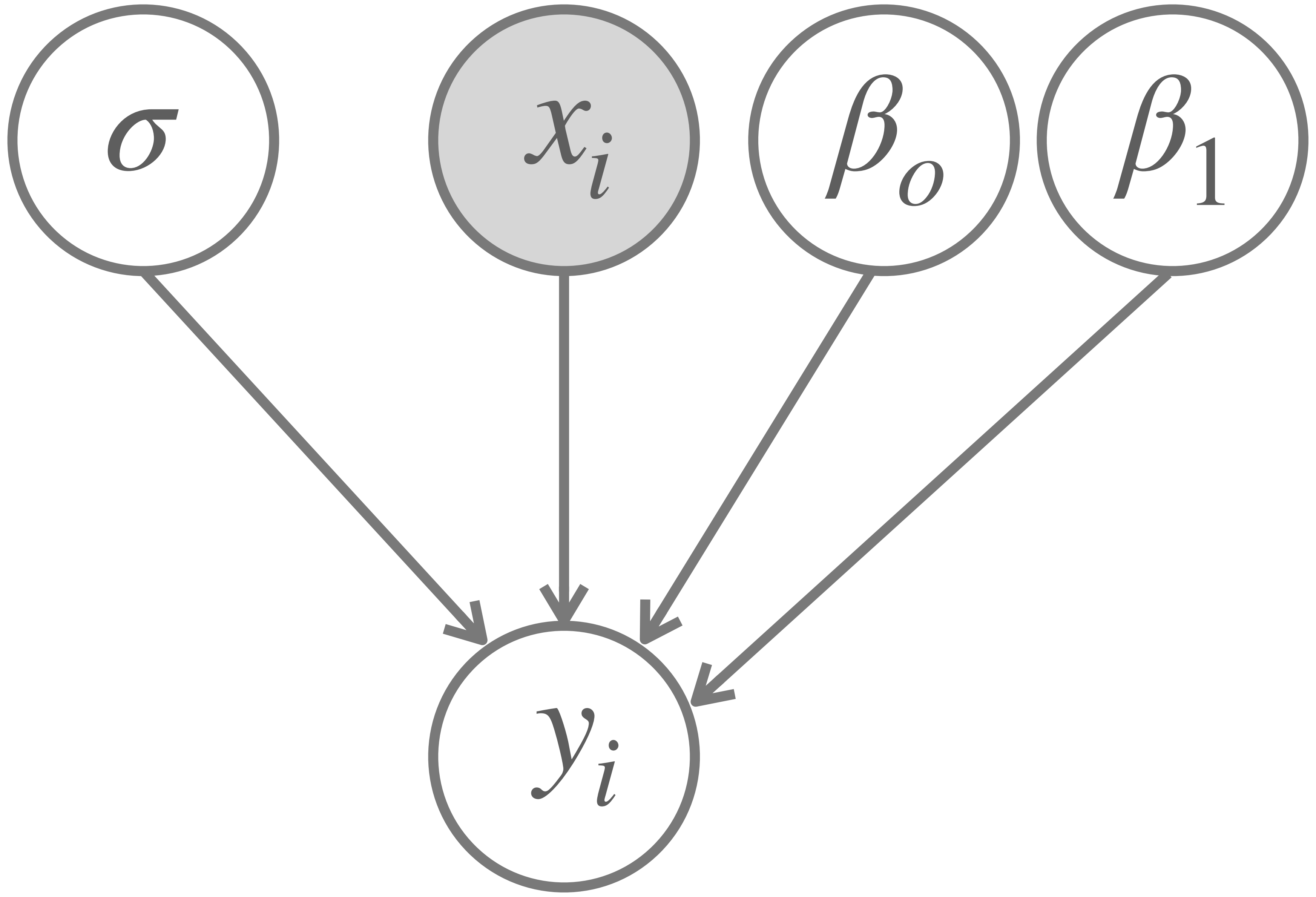
$$\sigma \sim \text{Trunc-Norm}(\ldots, \text{lower} = 0)$$

$$\mu \sim \text{Normal}(\ldots)$$

$$\delta \sim \text{Normal}(0, \ldots)$$

$$y_i \sim \begin{cases} \text{Normal}(\mu, \sigma) & \text{if } g_i = 0 \\ \text{Normal}(\mu + \delta, \sigma) & \text{if } g_i = 1 \end{cases}$$

SIMPLE LINEAR REGRESSION MODEL



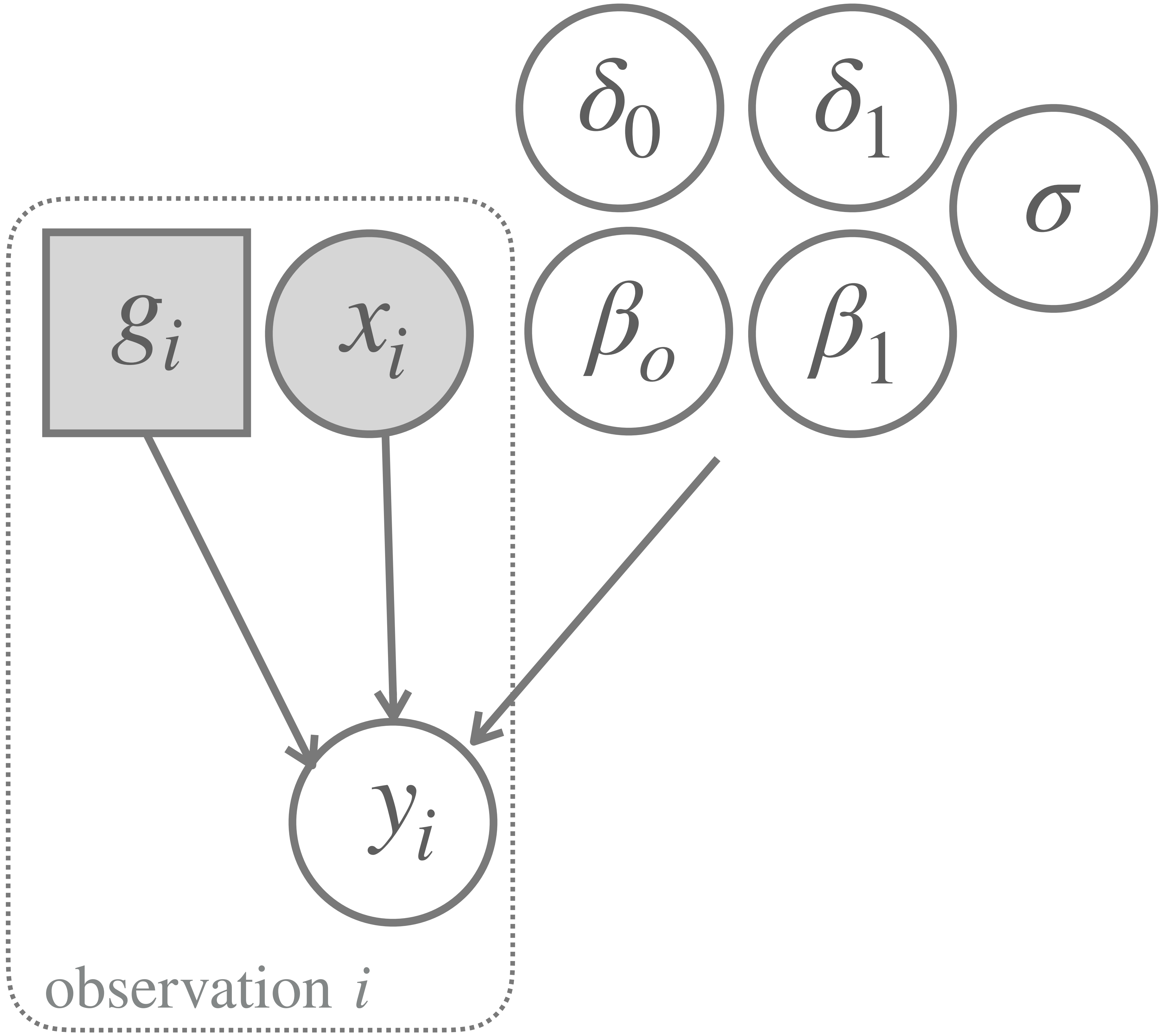
$$\sigma \sim \text{Trunc-Norm}(\ldots)$$

$$\beta_0 \sim \text{Student-t}(\ldots)$$

$$\beta_1 \sim \text{Student-t}(\ldots)$$

$$y_i \sim \text{Normal}(\beta_0 + \beta_1 x_i, \sigma)$$

LINEAR REGRESSION WITH TWO GROUPS



$$\sigma \sim \text{Trunc-Norm}(\ldots)$$

$$\delta_0 \sim \text{Student-T}(\mu = 0, \ldots)$$

$$\delta_1 \sim \text{Student-T}(\mu = 0, \ldots)$$

$$\beta_0 \sim \text{Student-t}(\ldots)$$

$$\beta_1 \sim \text{Student-t}(\ldots)$$

$$y_i \sim \begin{cases} \text{Normal}(\beta_0 + \beta_1 x_i, \sigma) & \text{if } g_i = 0 \\ \text{Normal}(\beta_0 + \delta_0 + (\beta_1 + \delta_0) x_i, \sigma) & \text{if } g_i = 1 \end{cases}$$