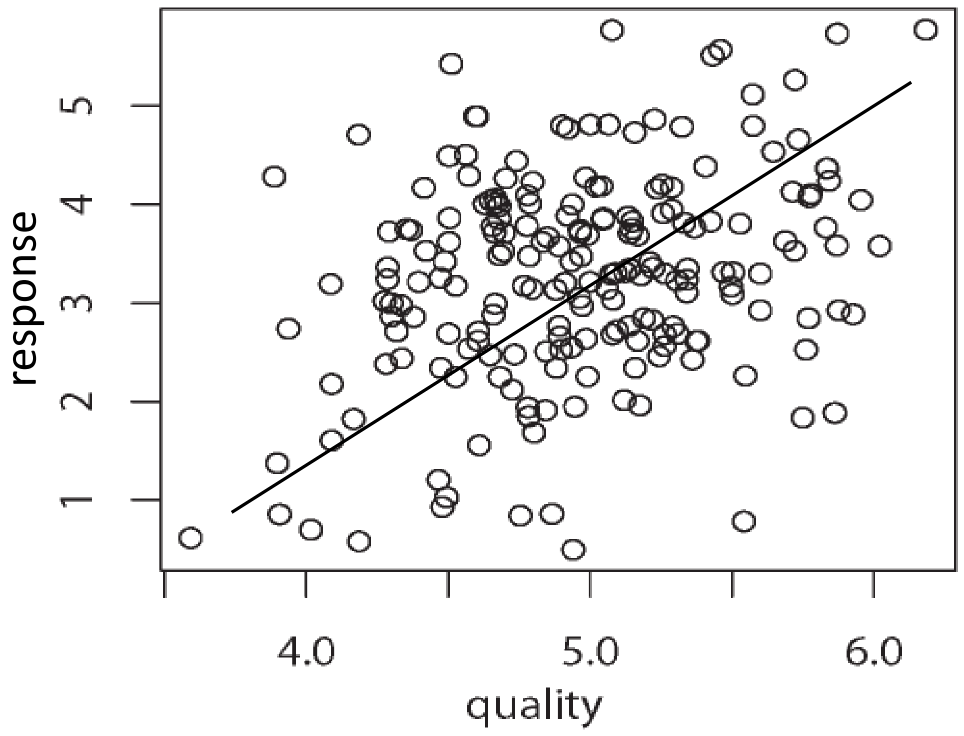
Characterizing Uncertainty

Classic Assumptions of Linear Model:

- Homoskedasticity
- No error in X variables
- Error in Y is measurement error
- Normally distributed error
- Observations are independent
- No missing data

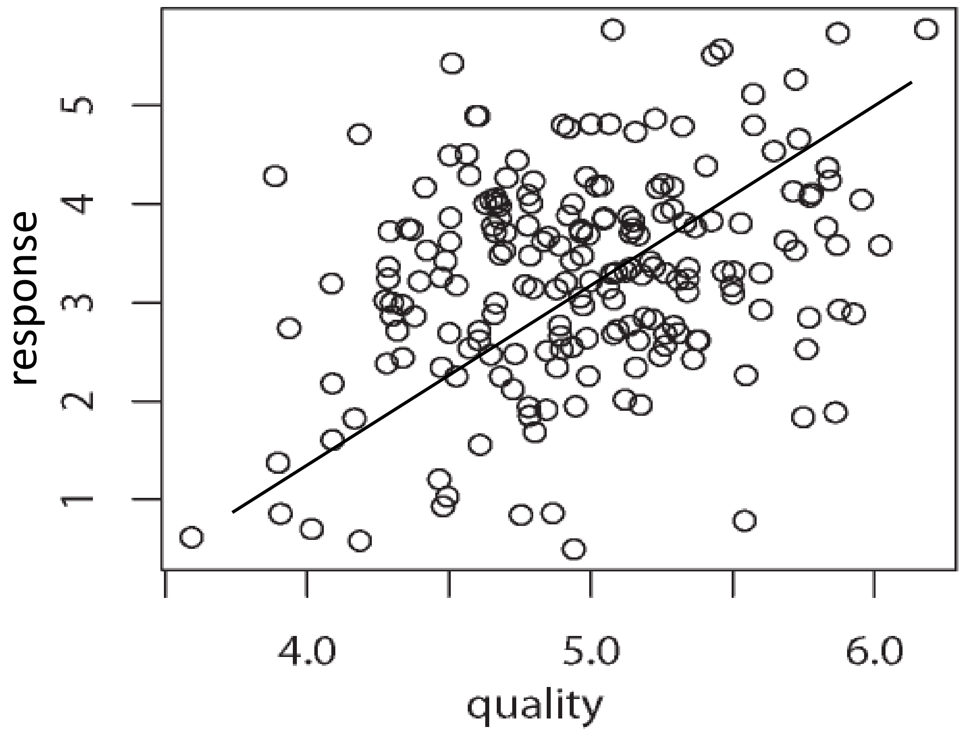
Classic Assumptions of Linear Model:

- Homoskedasticity
- No error in X variables
- Error in Y is measurement error
- Normally distributed error
- Observations are independent
- No missing data



Classic Assumptions:

$$y_i \sim \beta_0 + \beta(x_i) + \varepsilon_i$$

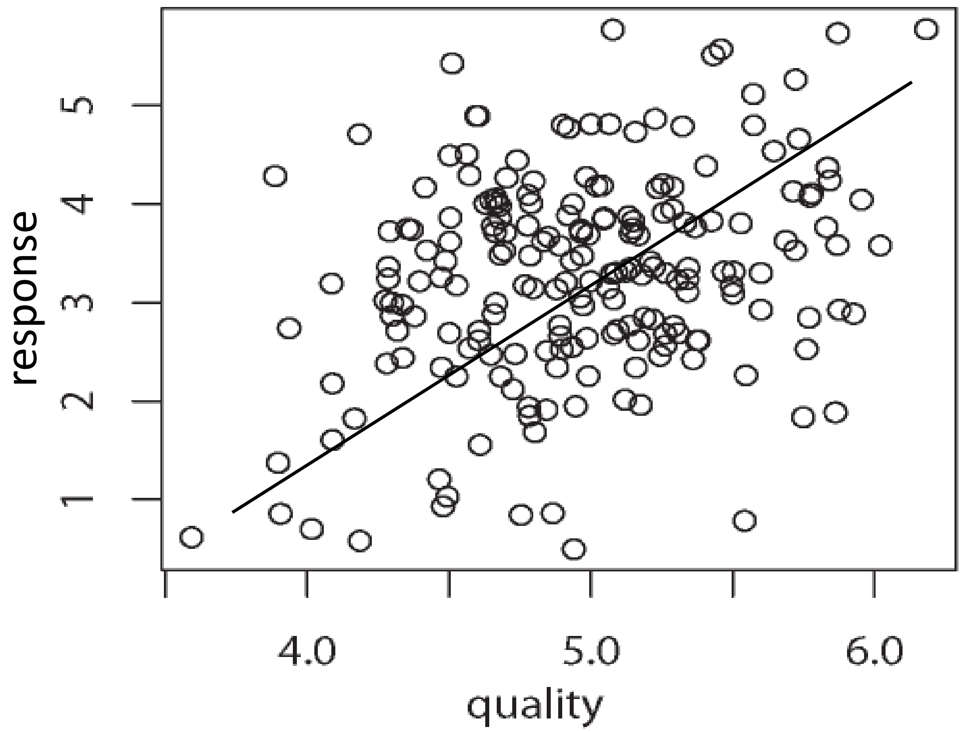


Classic Assumptions:

$$y_i \sim \beta_0 + \beta(x_i) + \varepsilon_i$$

Data Model $y_i \sim N(\mu_i, \varepsilon_i)$

Process Model $\mu_i = \beta_0 + \beta_1(x_i)$



Classic Assumptions:

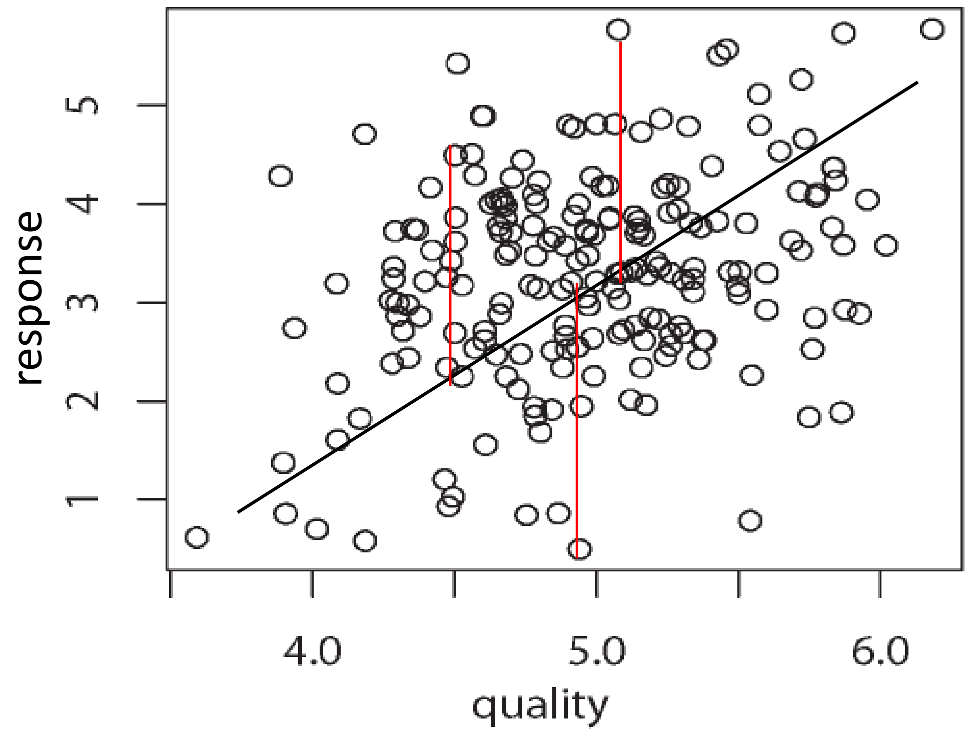
$$y_i \sim \beta_0 + \beta(x_i) + \varepsilon_i$$

Data Model $y_i \sim N(\mu_i, \varepsilon_i)$

Process Model $\mu_i = \beta_0 + \beta_1(x_i)$

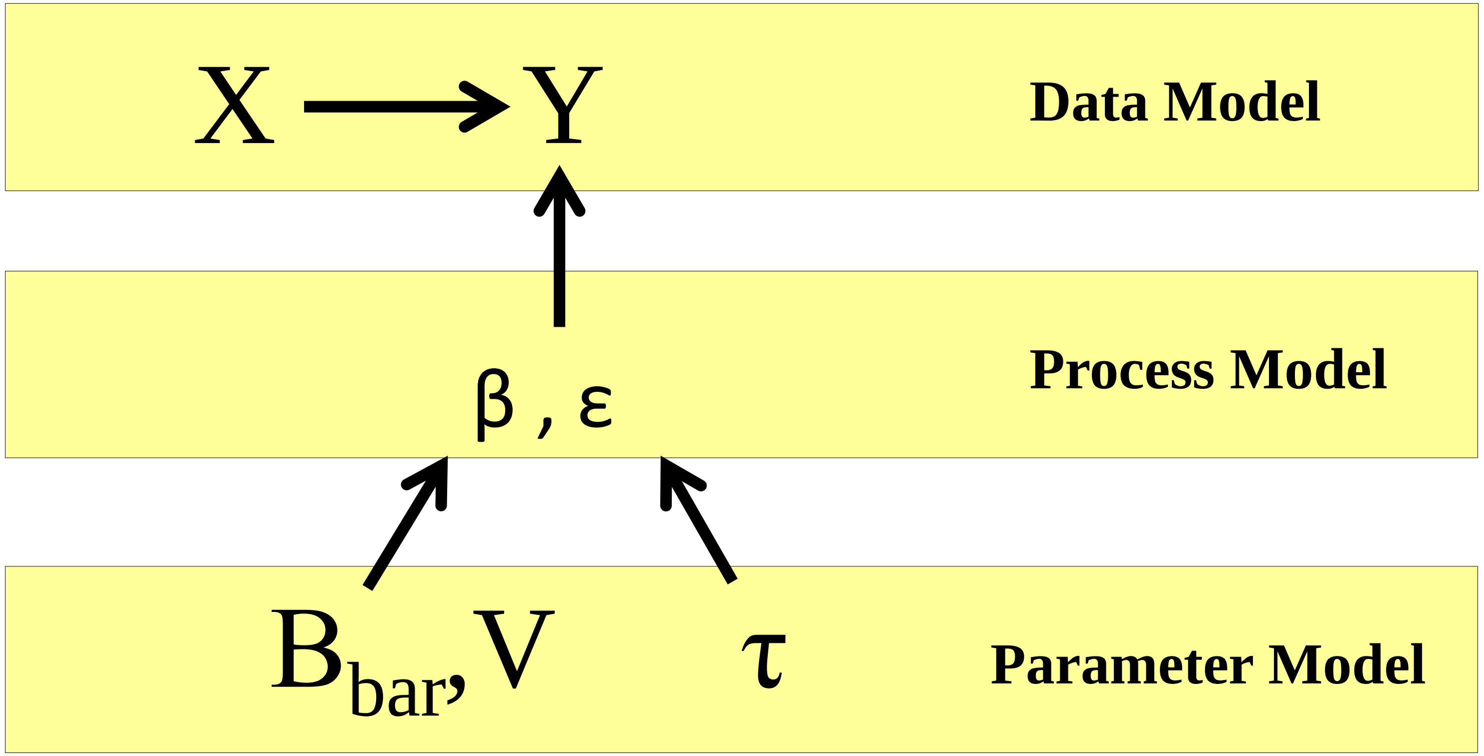
Parameter Model $\varepsilon_i \sim N(0,\tau)$

$\beta \sim N(\beta_{bar}, \nu)$



Linear Model – Graph Notation

$$y_i \sim \beta_0 + \beta(x_i) + \varepsilon_i$$



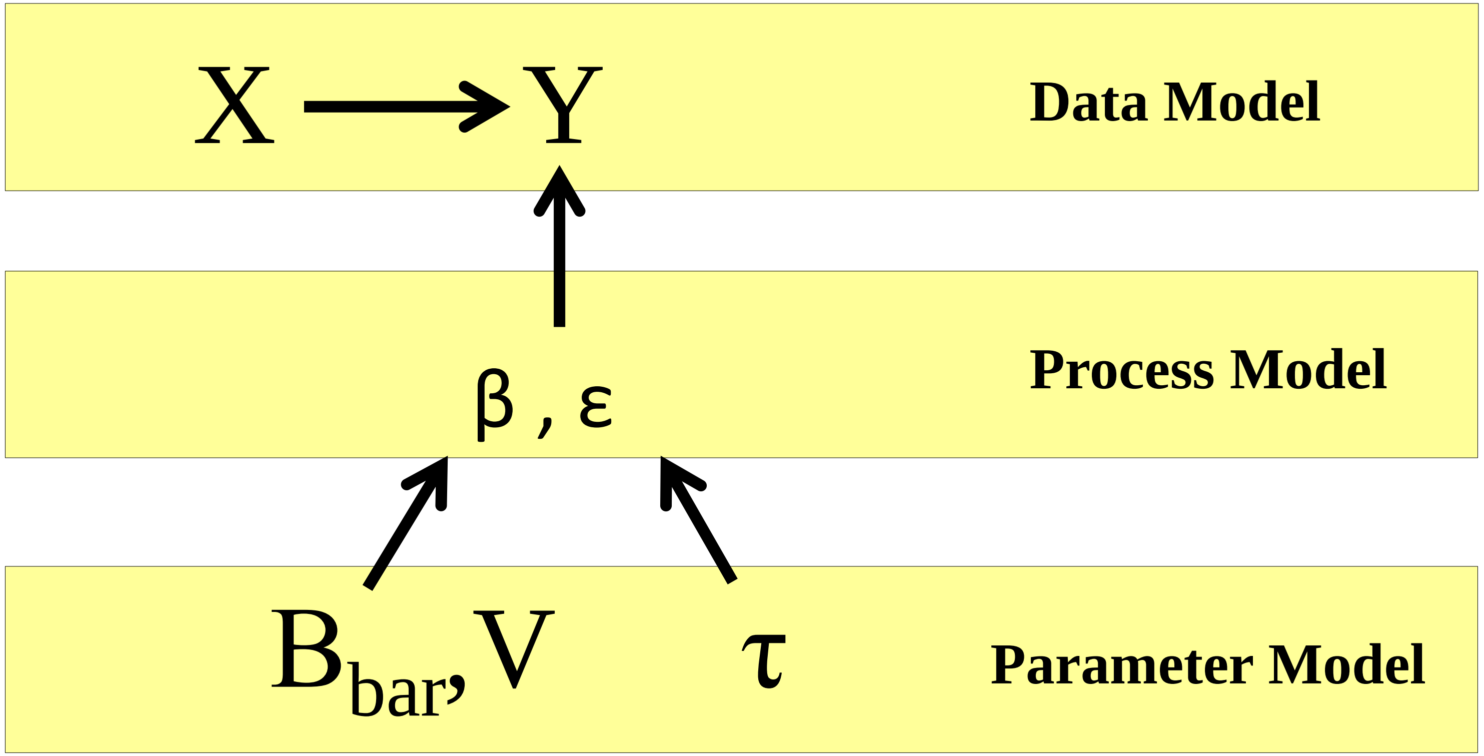
Linear Model – Graph Notation

$y_i \sim N(\mu_i, \varepsilon_i)$

$\mu_i = \beta_0 + \beta_1(x_i)$

$\varepsilon_i \sim N(0,\tau)$

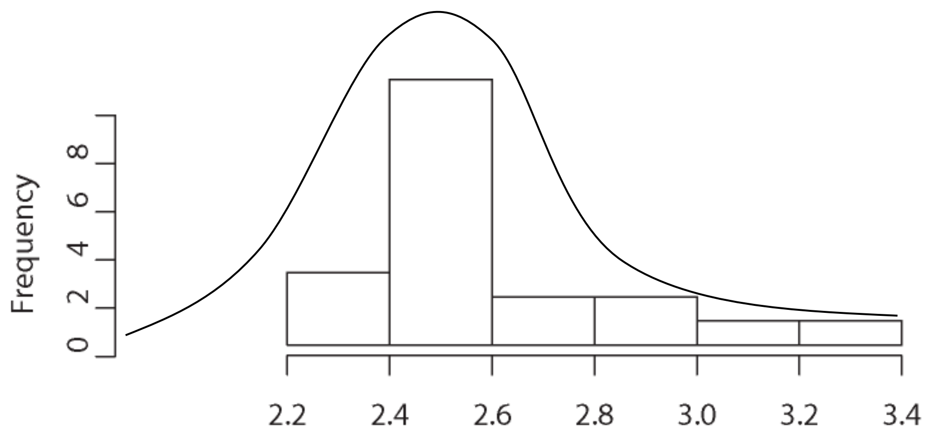
$\beta \sim N(\beta_{bar}, \nu)$



Beyond the classic assumptions

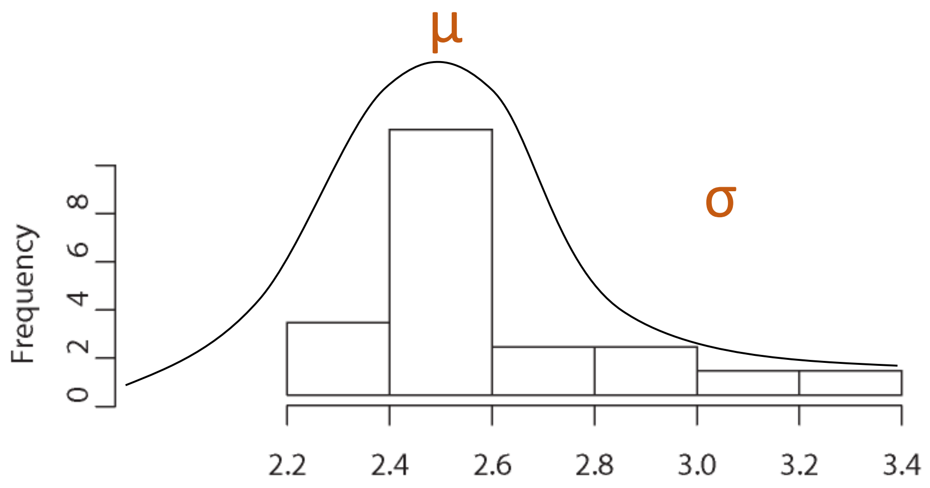
(or what to do with real ecological data)

We assume data (bars) are random samples from the true population (line). The expected relationship between the samples and the true population is described by a probability distribution.



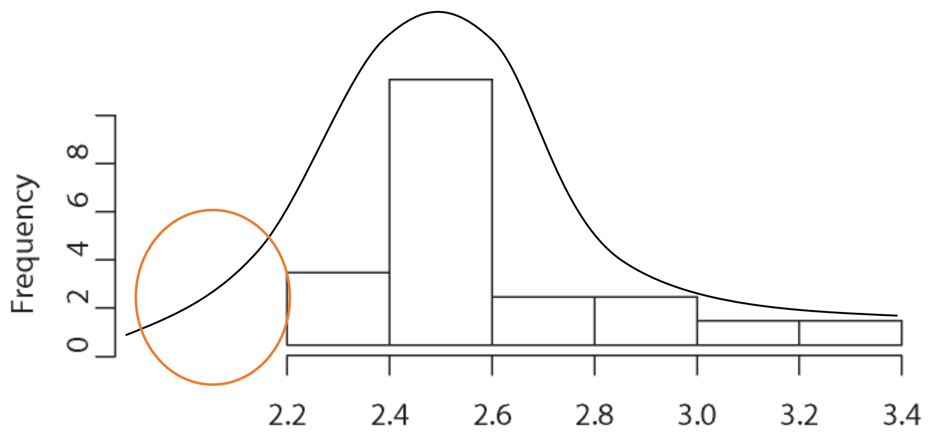
$$p(y|\mu) \propto \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

We assume data (bars) are random samples from the true population (line). The expected relationship between the samples and the true population is described by a probability distribution.



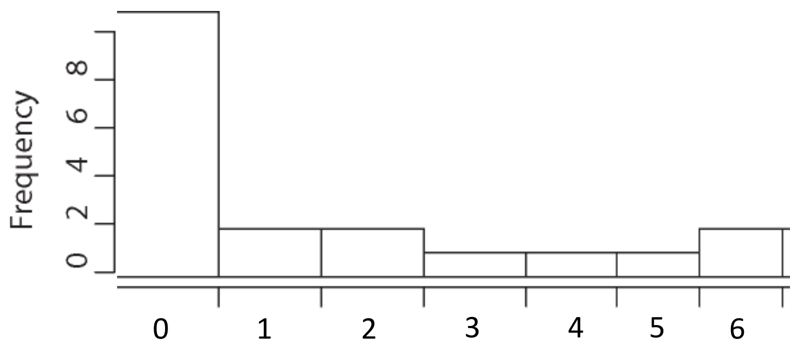
$$p(y|\mu) \propto \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

We assume data (bars) are random samples from the true population (line). The expected relationship between the samples and the true population is described by a probability distribution.

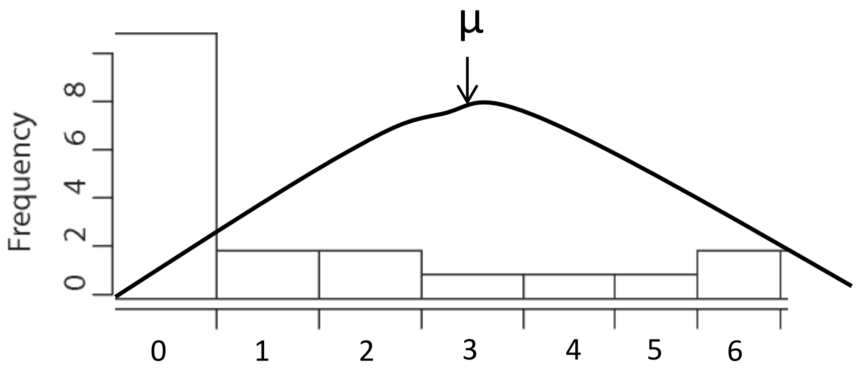


$$p(y|\mu) \propto \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

These data don’t look normal

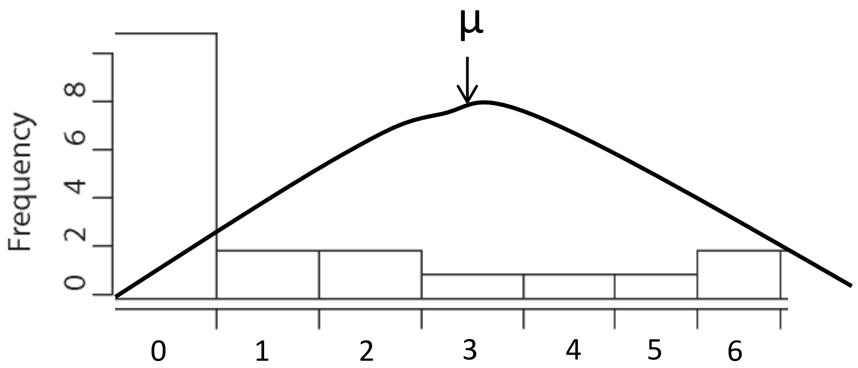


These data don't look normal

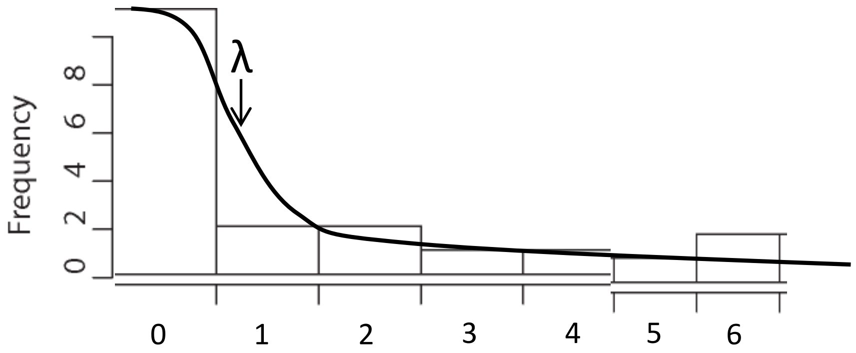


$$\propto exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

These data don't look normal

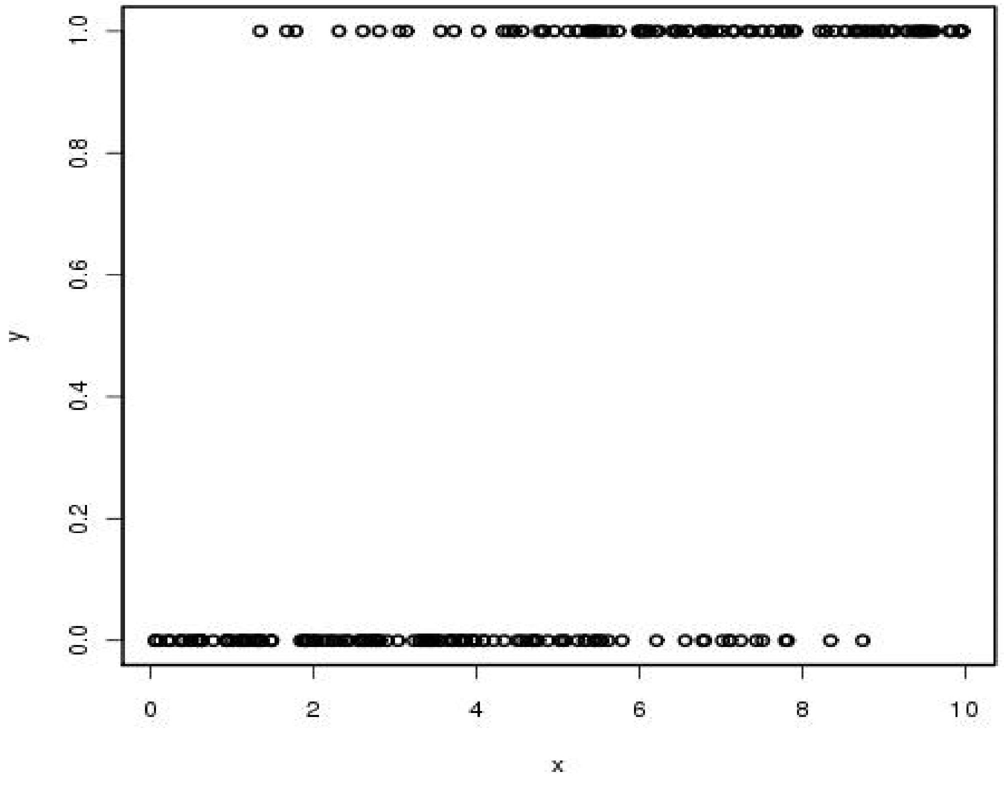


$$\propto exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

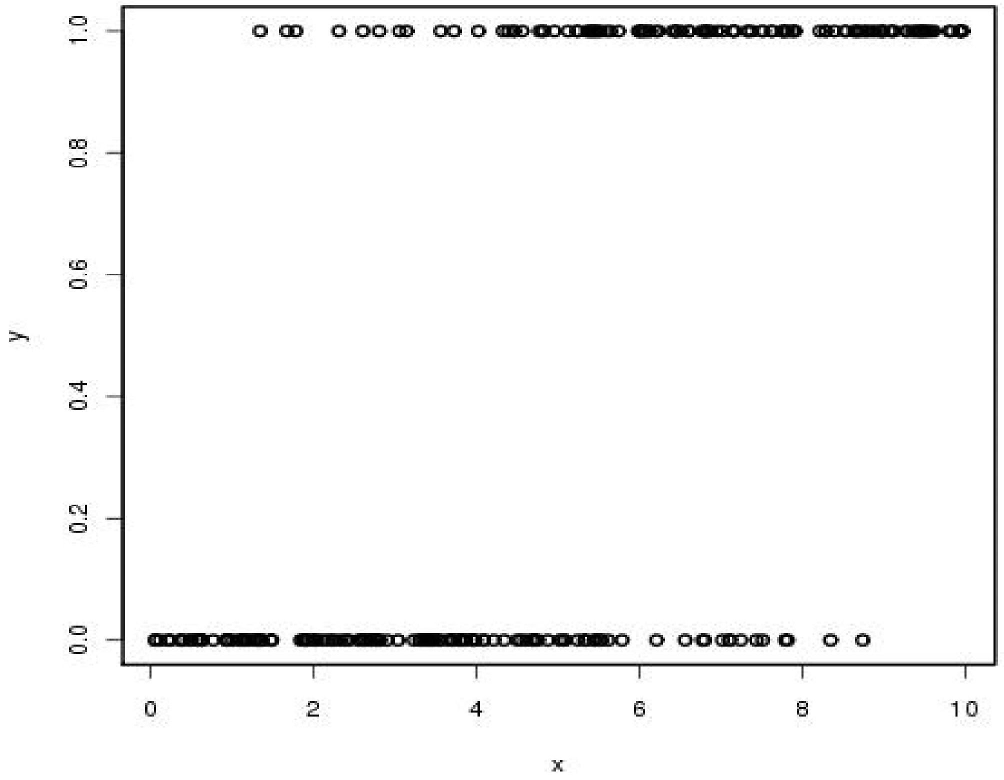


$$\frac{\lambda^x}{x!}\ e^{-\lambda}$$

These data don’t look normal



These data don't look normal

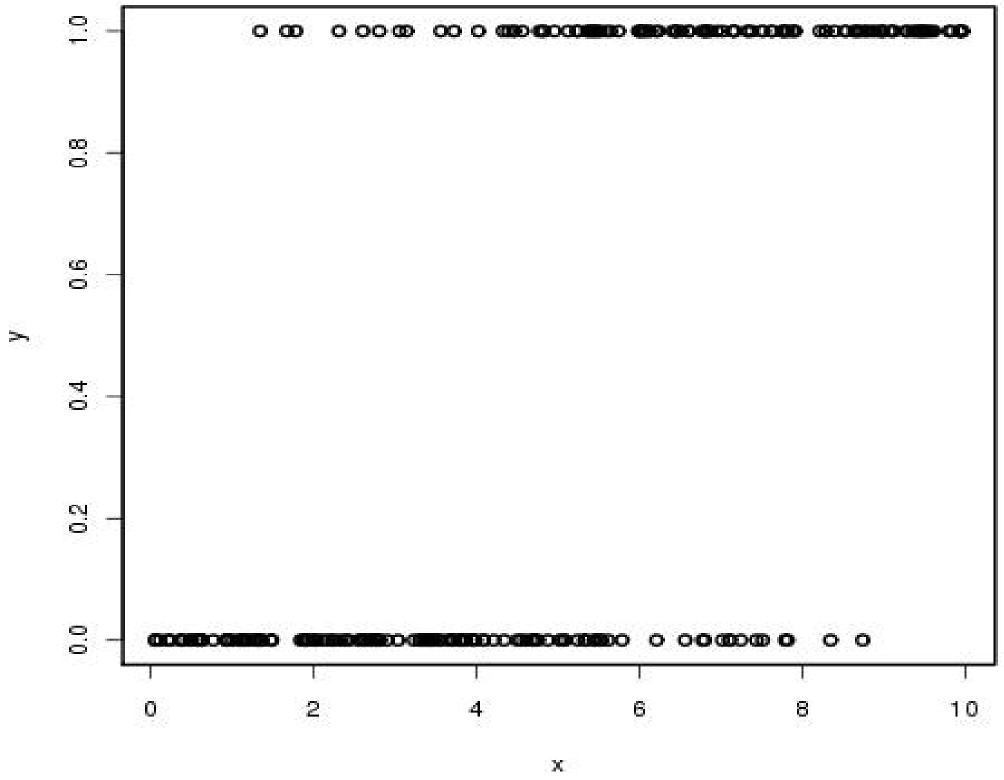


$$y_i \sim Bern(\rho_i)$$

$$\rho_i = logit\ (\theta_i) = \frac{1}{1 + e^{-\theta_i}}$$

$$\theta_i = \beta 0 + \beta x_i$$

These data don't look normal



$$y_i \sim Bern(\rho_i)$$ **Data Model**

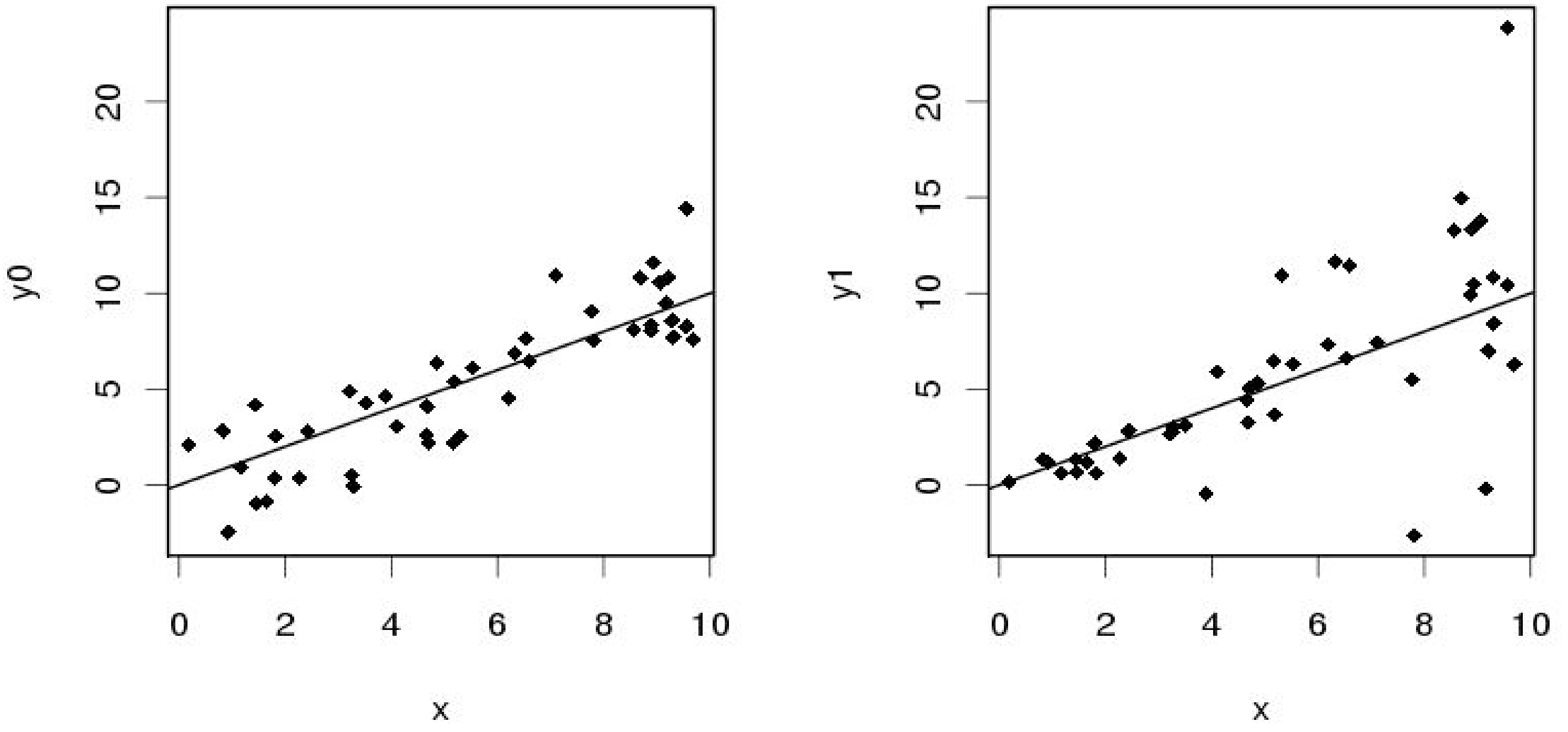
$$\rho_i = logit\ (\theta_i) = \frac{1}{1 + e^{-\theta_i}}$$ Link

$$\theta_i = \boldsymbol{\beta} X$$ Linear

Process Model

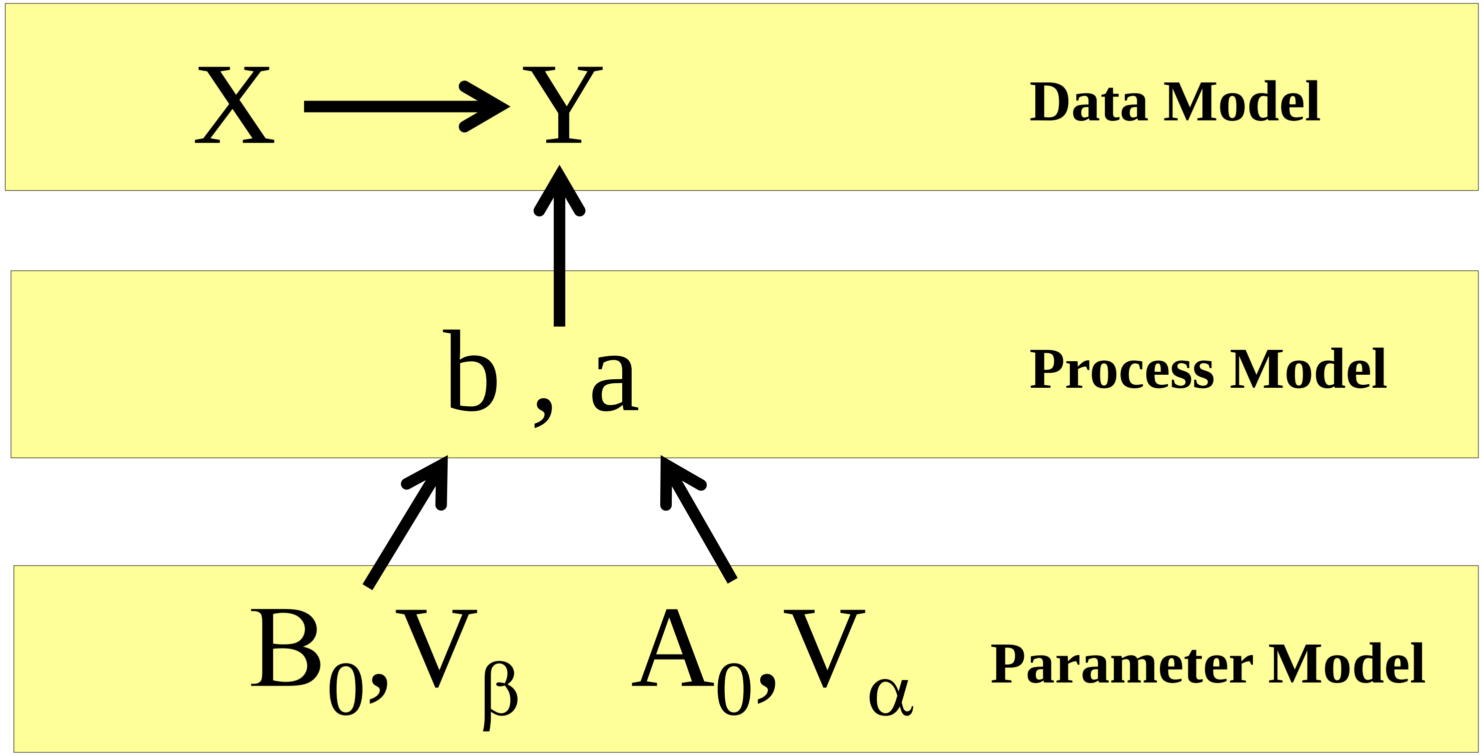
$$\boldsymbol{\beta} \sim Norm(\beta 0, \tau)$$ **Parameter Model**

Variance



Heteroskedasticity

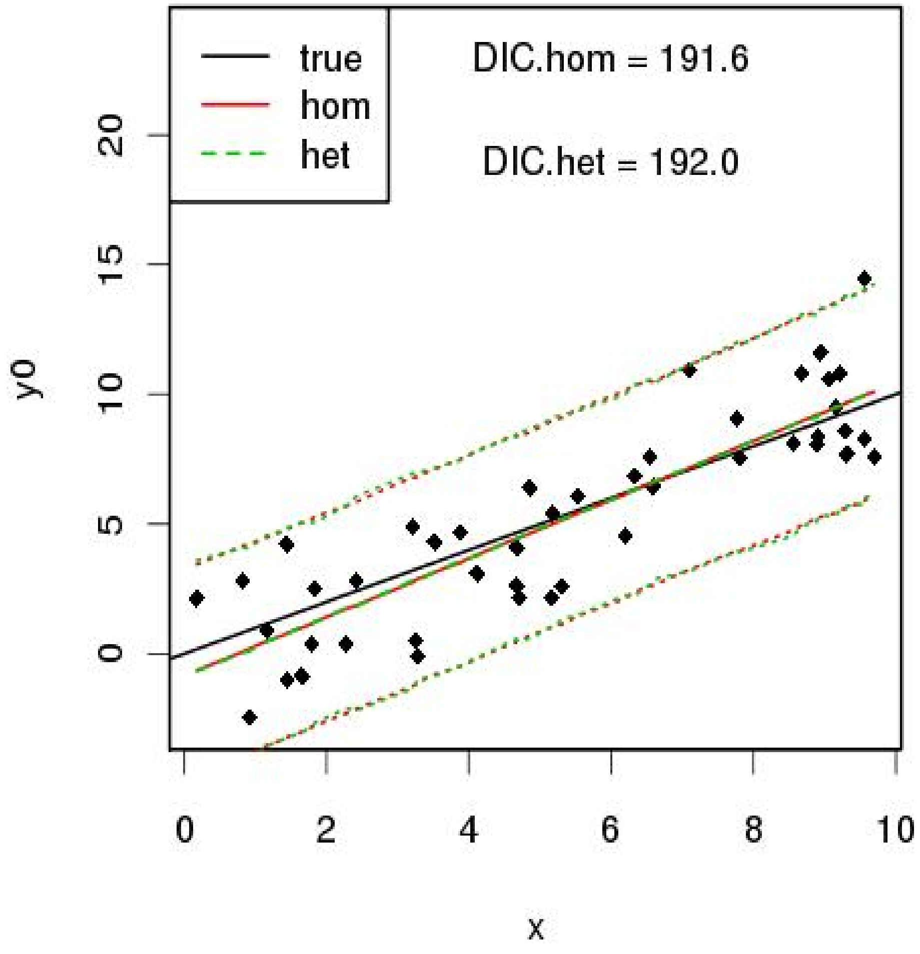
$$y \sim N\left(\beta_1 + \beta_2 x, (\alpha_1 + \alpha_2 x)^2\right)$$

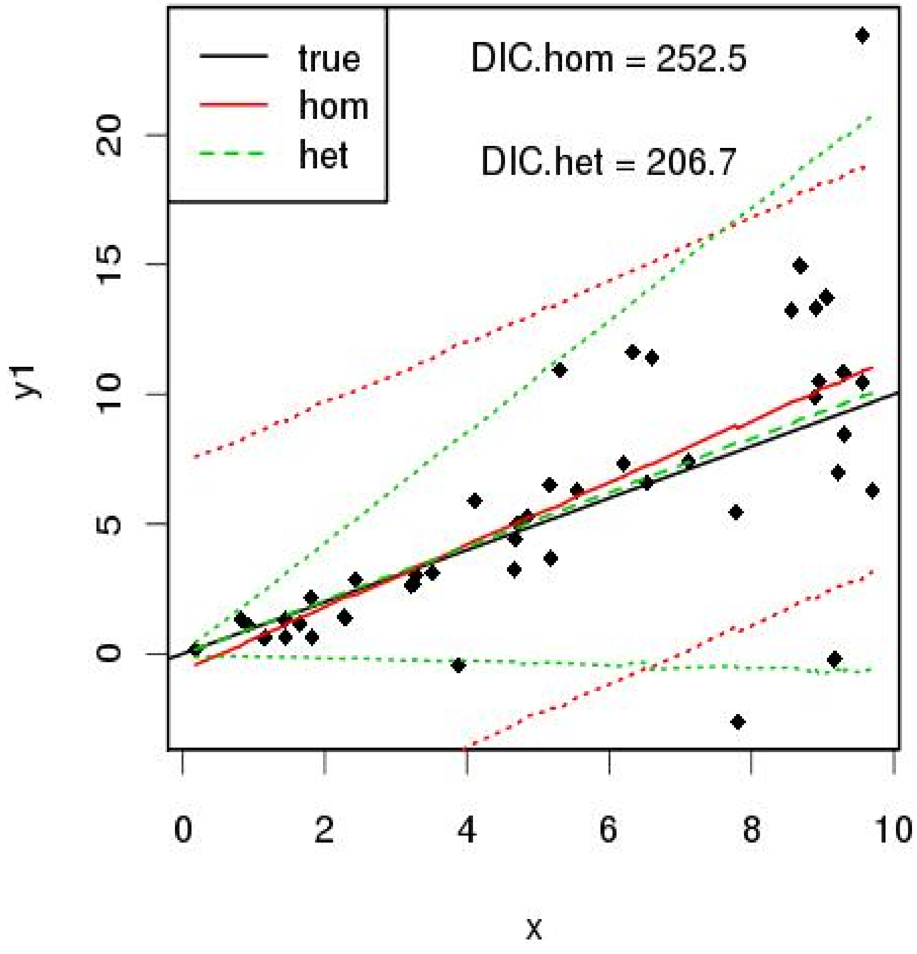


Heteroskedasticity

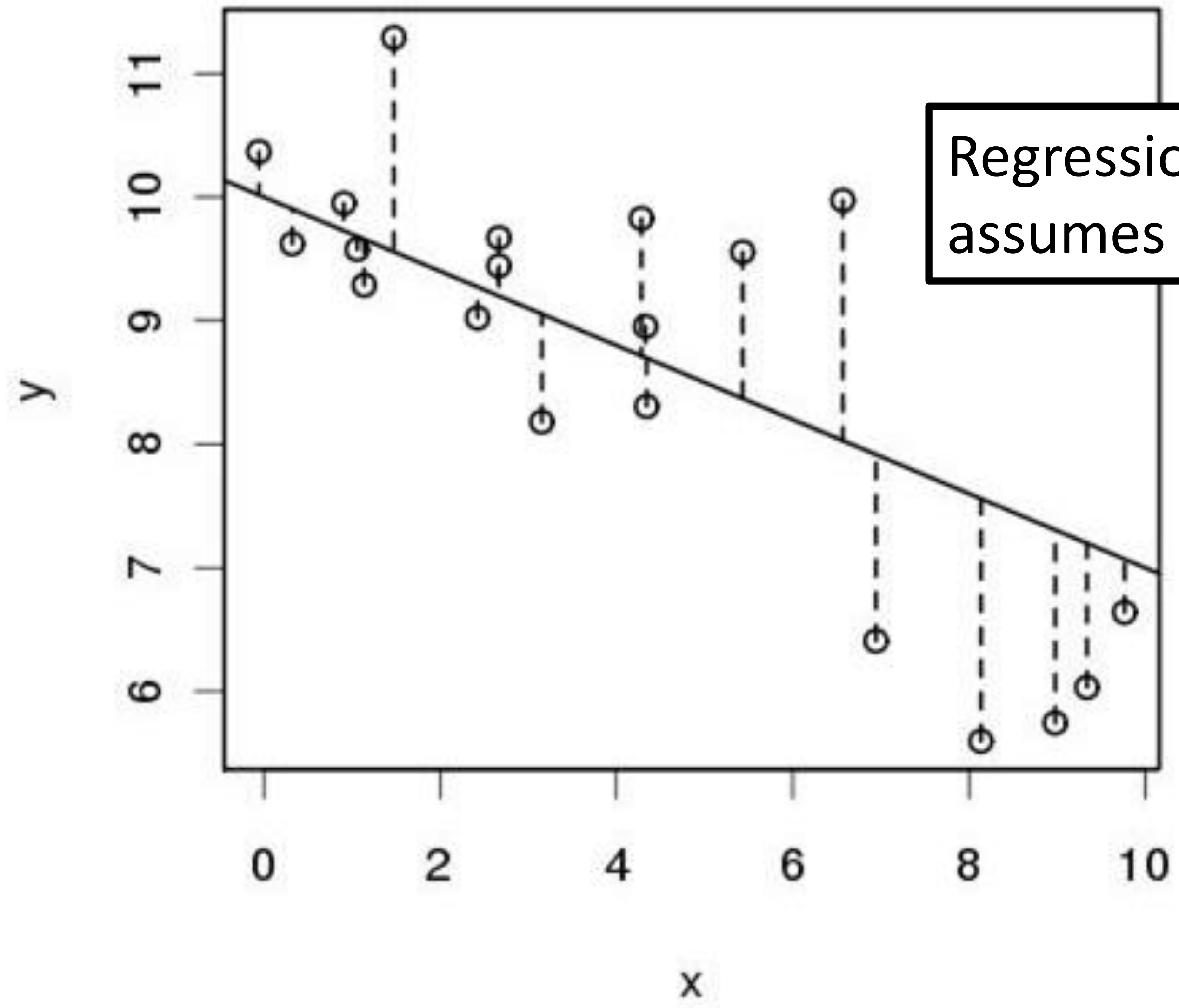
$$y \sim N\left(\beta_1 + \beta_2 x, s^2\right)$$

$$y \sim N\left(\beta_1 + \beta_2 x, (\alpha_1 + \alpha_2 x)^2\right)$$



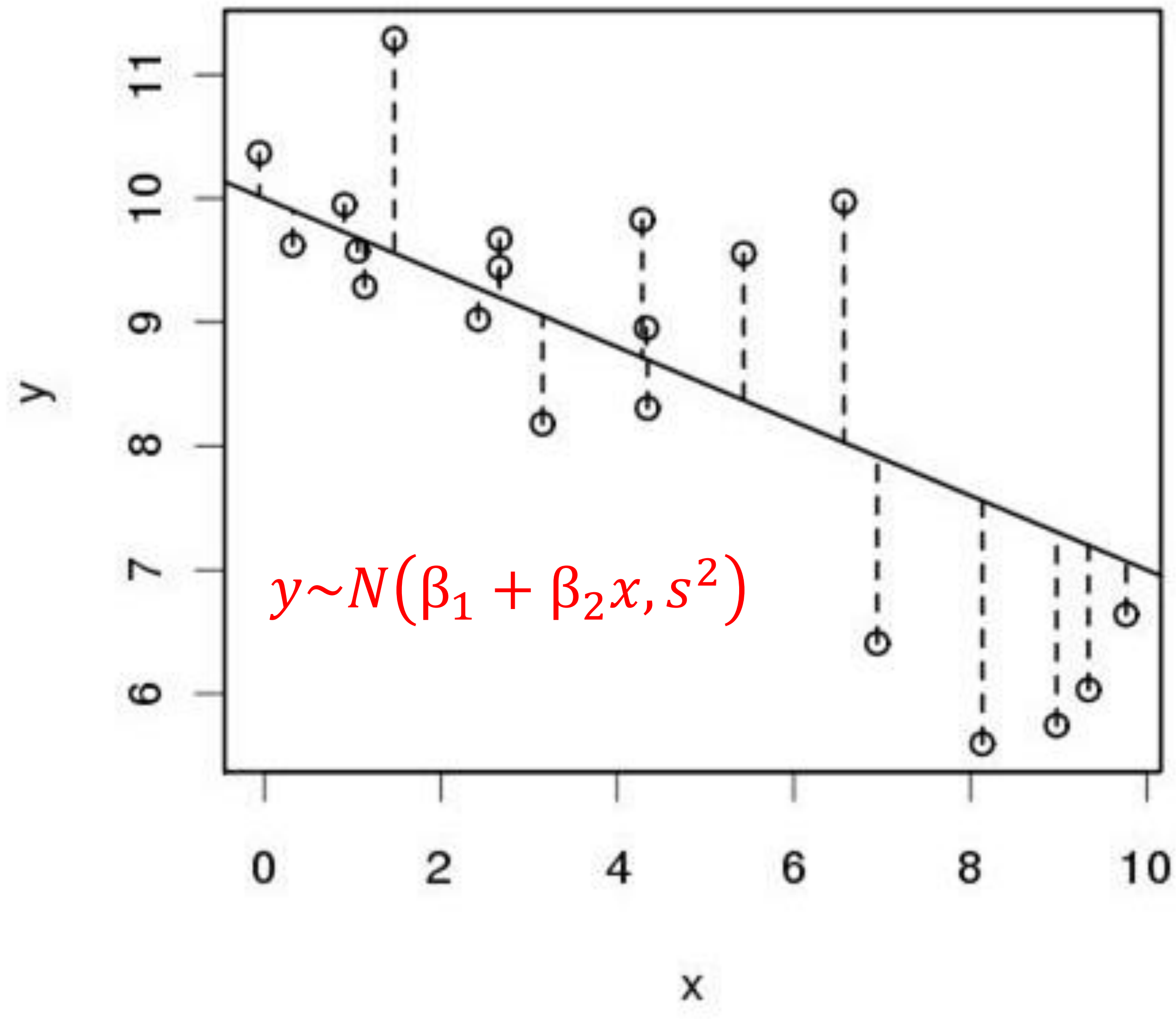


Observation error



Regression model assumes all error is in Y

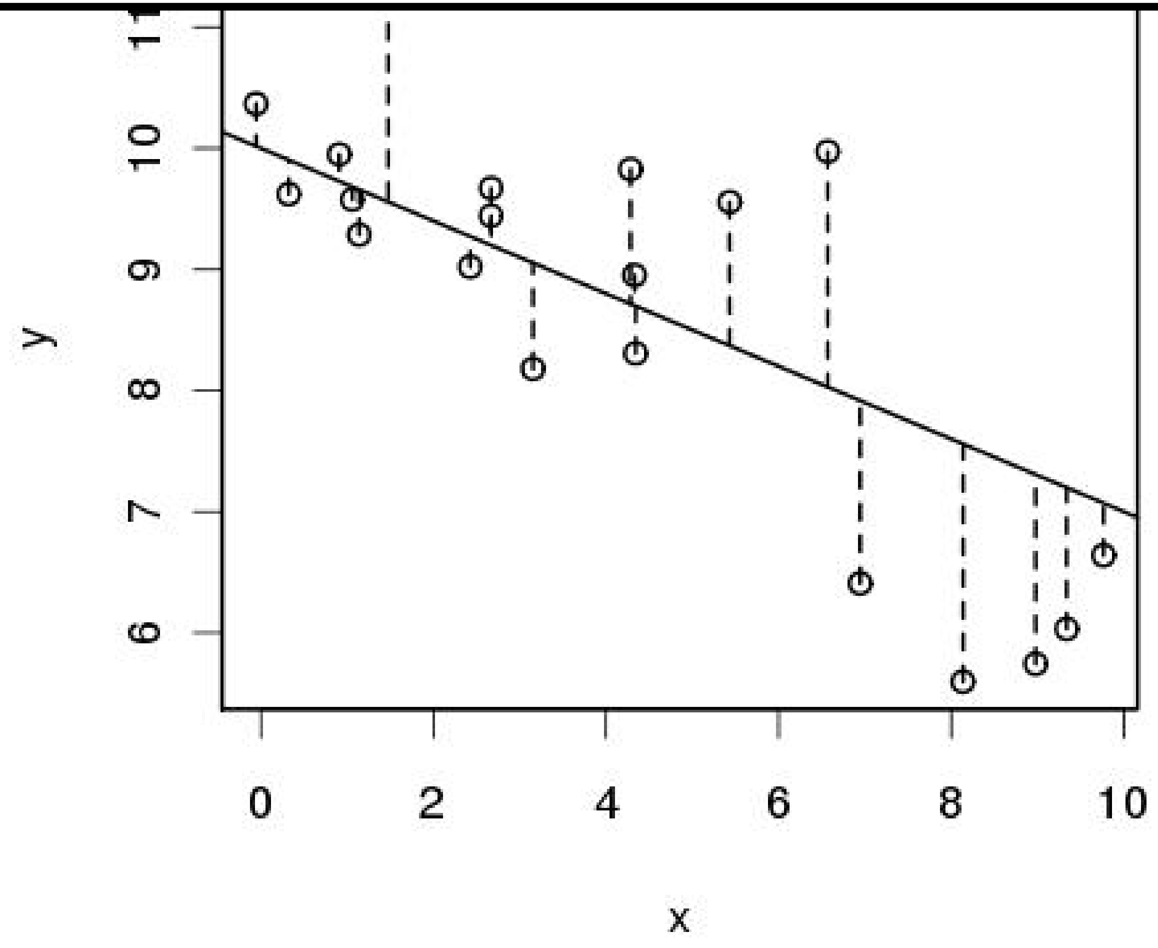
Observation error



Errors in variables

But we're not always great at measuring x either…

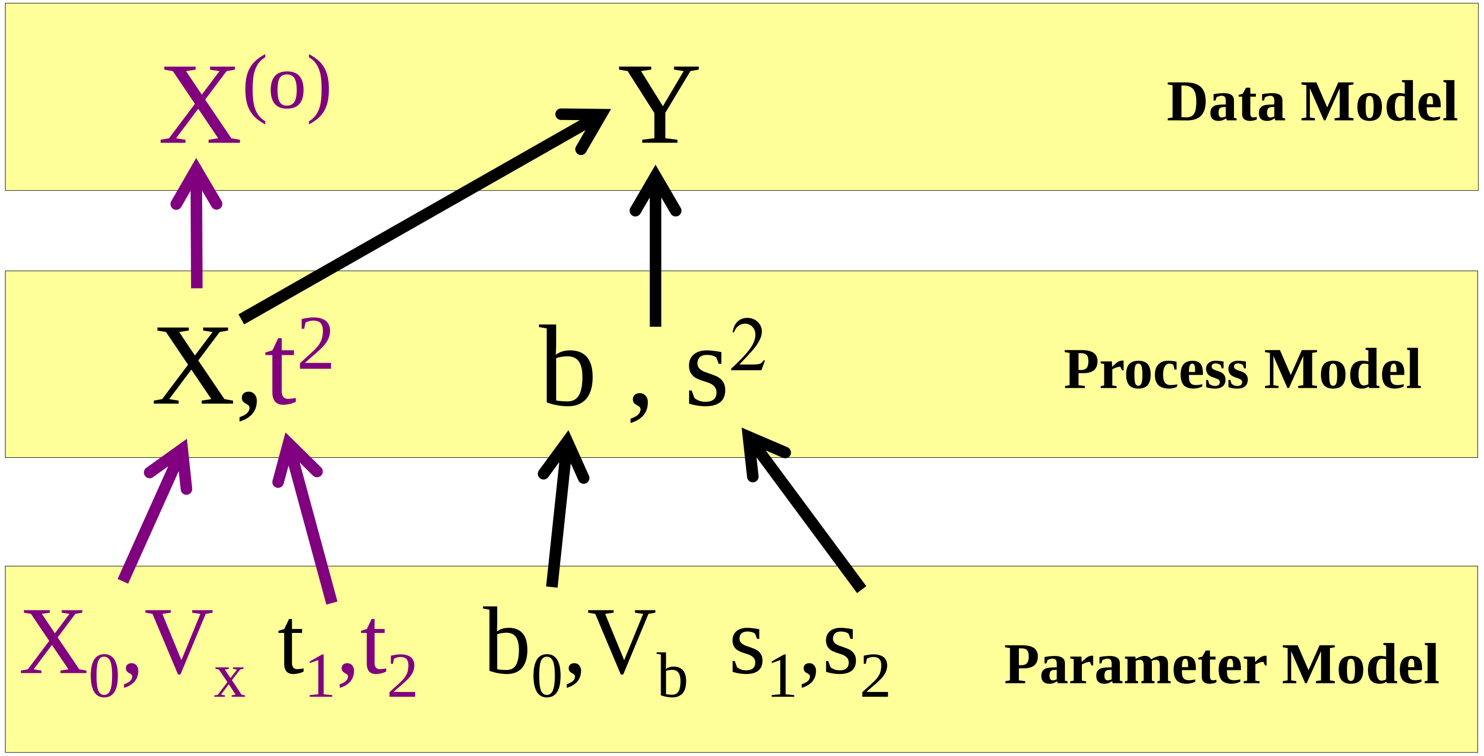
$8 = \beta_1 + \beta_2(?)$



Errors in variables

$$y \sim N\left(\beta_1 + \beta_2 x, s^2\right)$$

$$x^{(o)} \sim N\left(x, \tau^2\right)$$ Model x as random variable

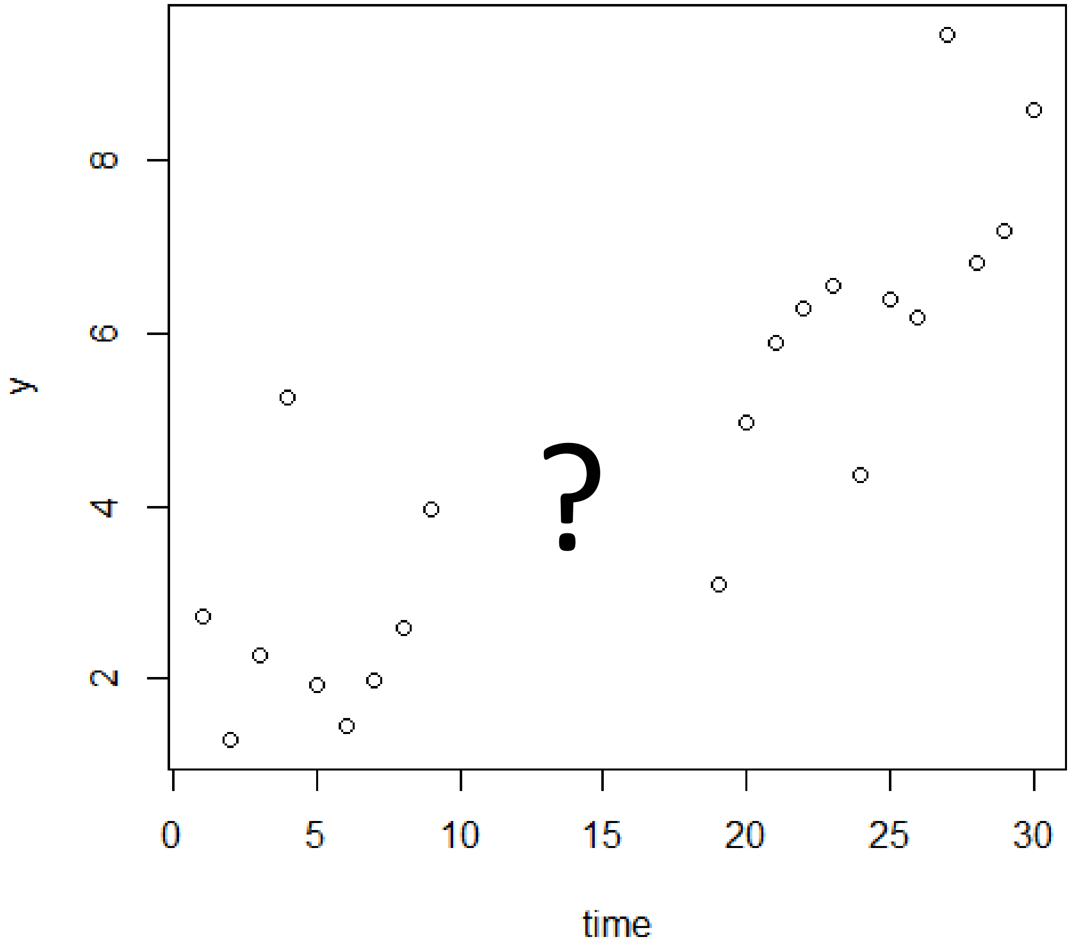


Latent Variables

- Any variable not directly observed
  - Missing data
  - Variable measured with error (bias or random)
  - Proxy measures

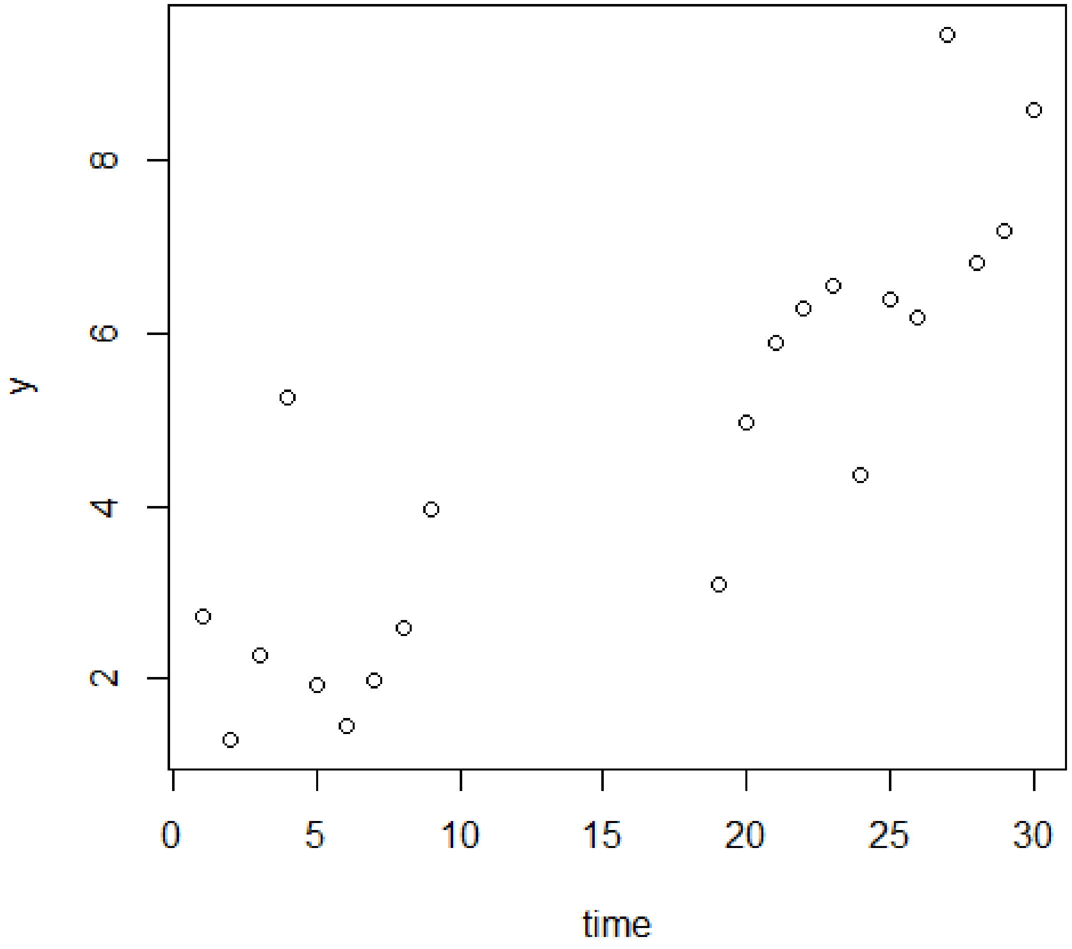
- Ignoring variable latency (e.g., modeling a derived response or flawed observation) can lead to incorrect or falsely <u>overconfident</u> conclusions

Missing Data



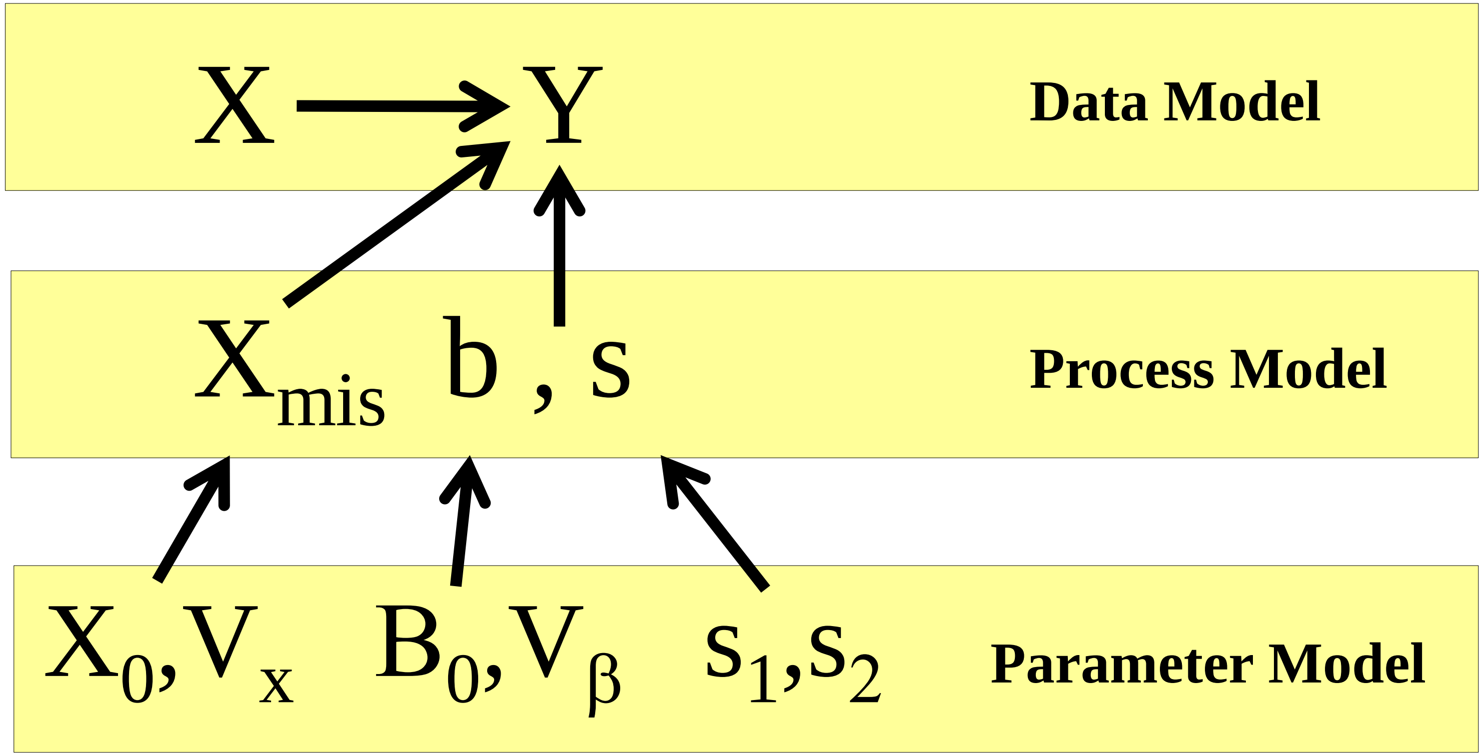
Missing Data

$$y* \sim N\left(\beta X, s^2\right)$$



Missing Data Model

$$y \sim N(\beta X, s^2)$$



Missing Data Model

- Update the regression model based on ALL the rows of data conditioned on the current values of the missing data
- Update the missing data based on the current regression model and the values that all other covariates take on

ASSUMPTION!!

- Missing data models assume that the data is *missing at random*
- If data is missing SYSTEMATICALLY it can not be estimated

Latent Variables

When you observe y but interpret as z…

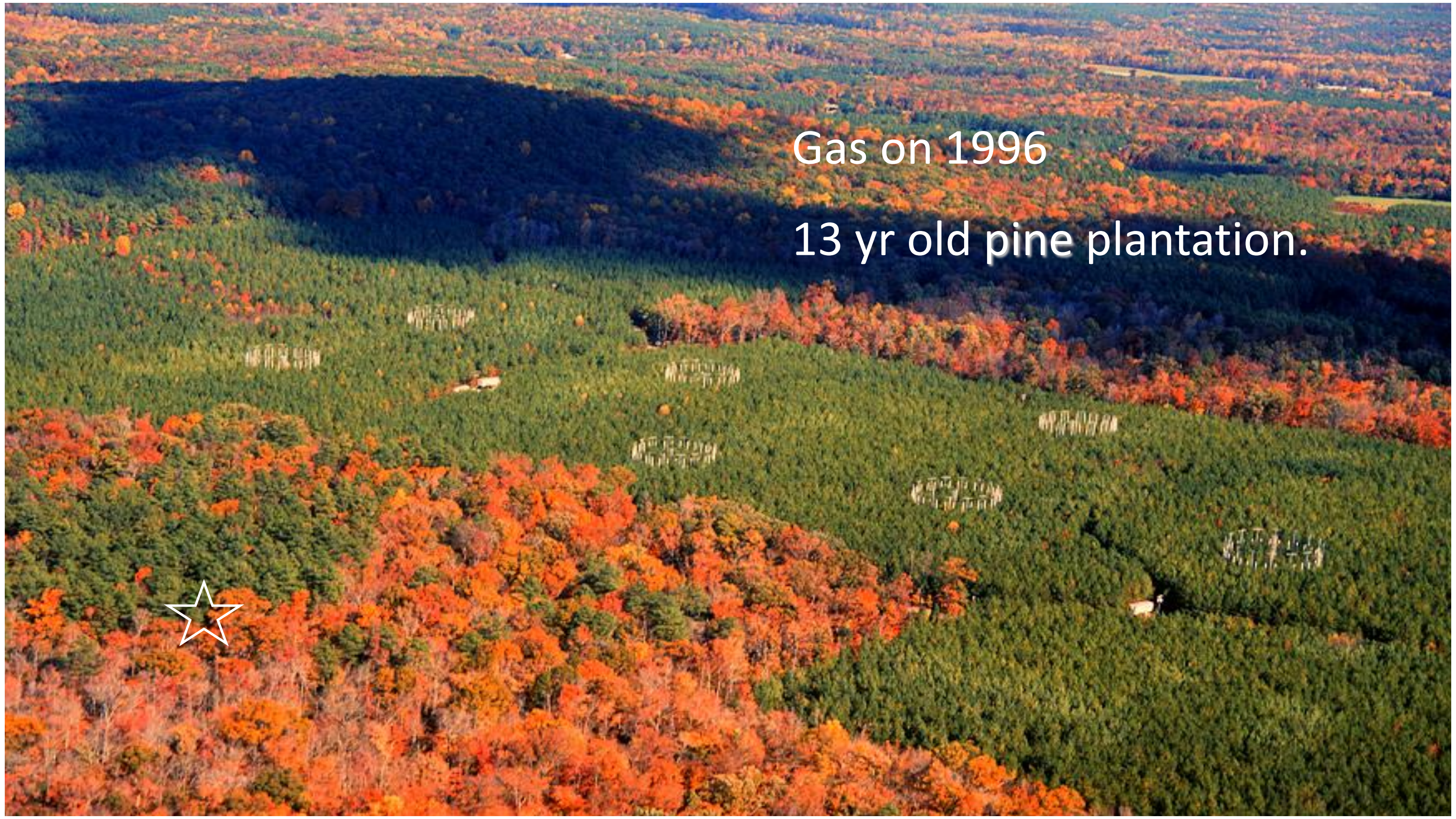
- observation error (random or biased)
- missing data
- proxies- one or multiple

Free Air Carbon Enrichment (FACE)

30 meters

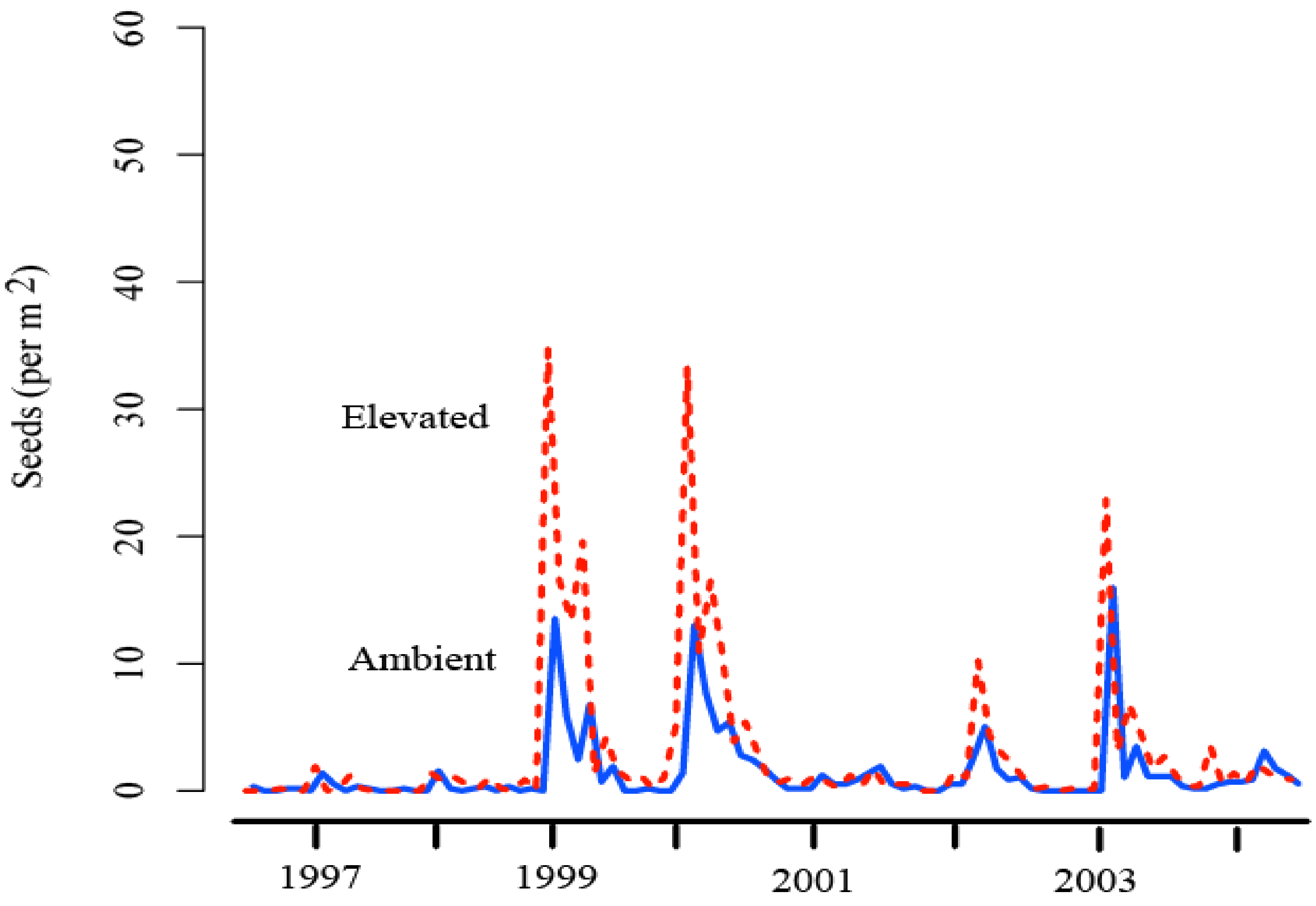
3 control rings: 365 ppm

3 treatment rings: 565 ppm



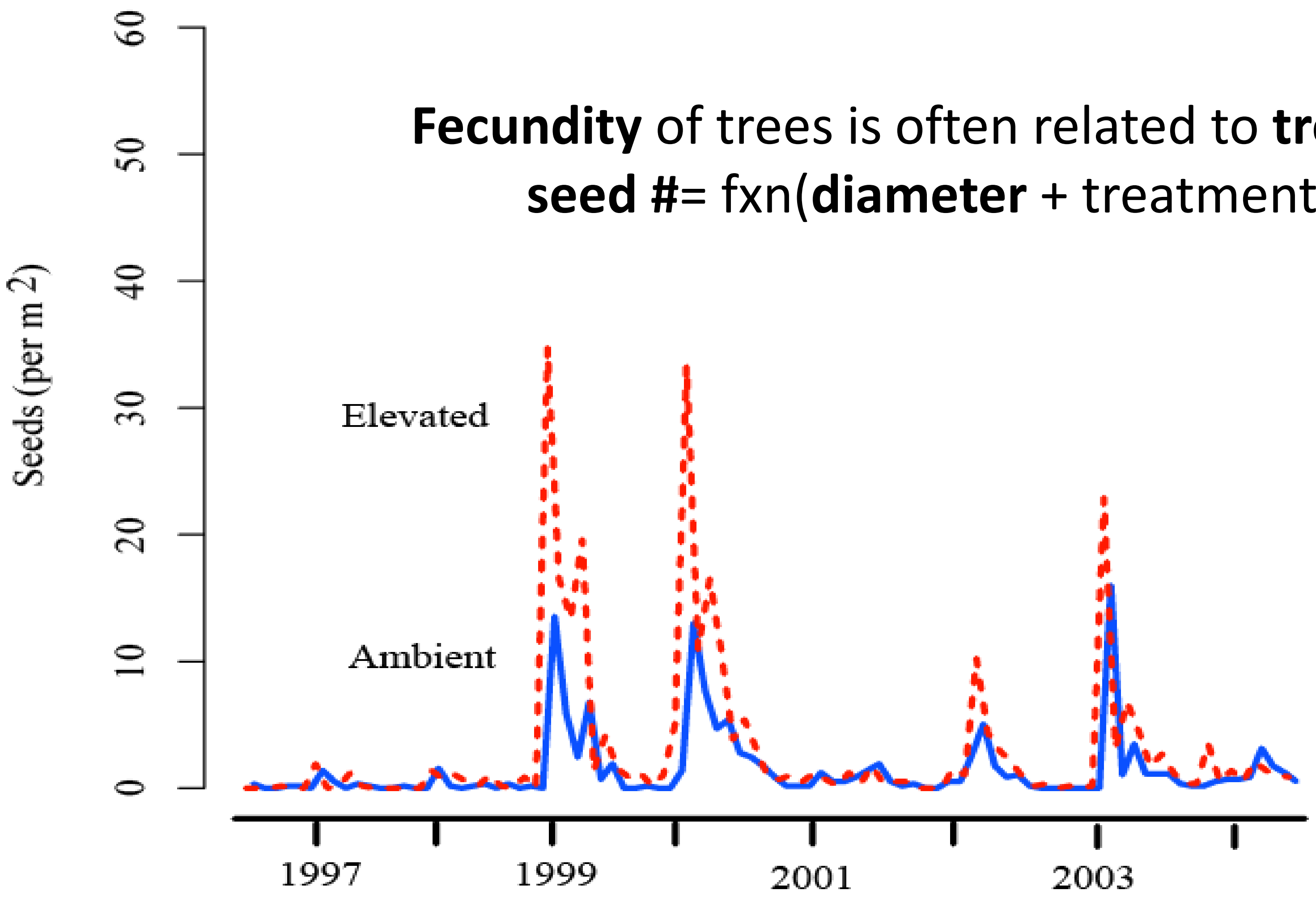
http://cdiac.edd.ornl.gov/programs/FACE/

CO_2 Fumigation Response

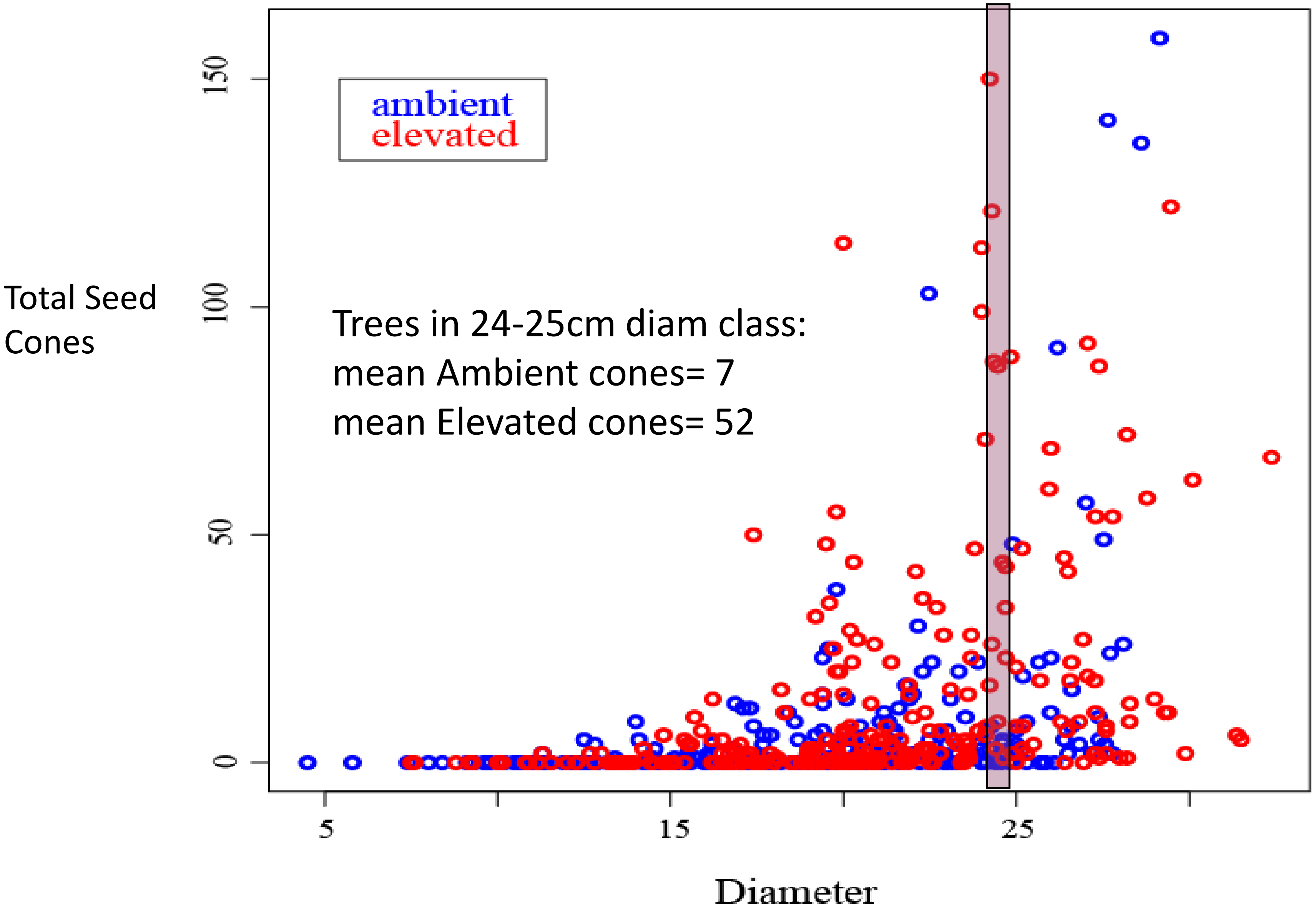


CO_2 Fumigation Response

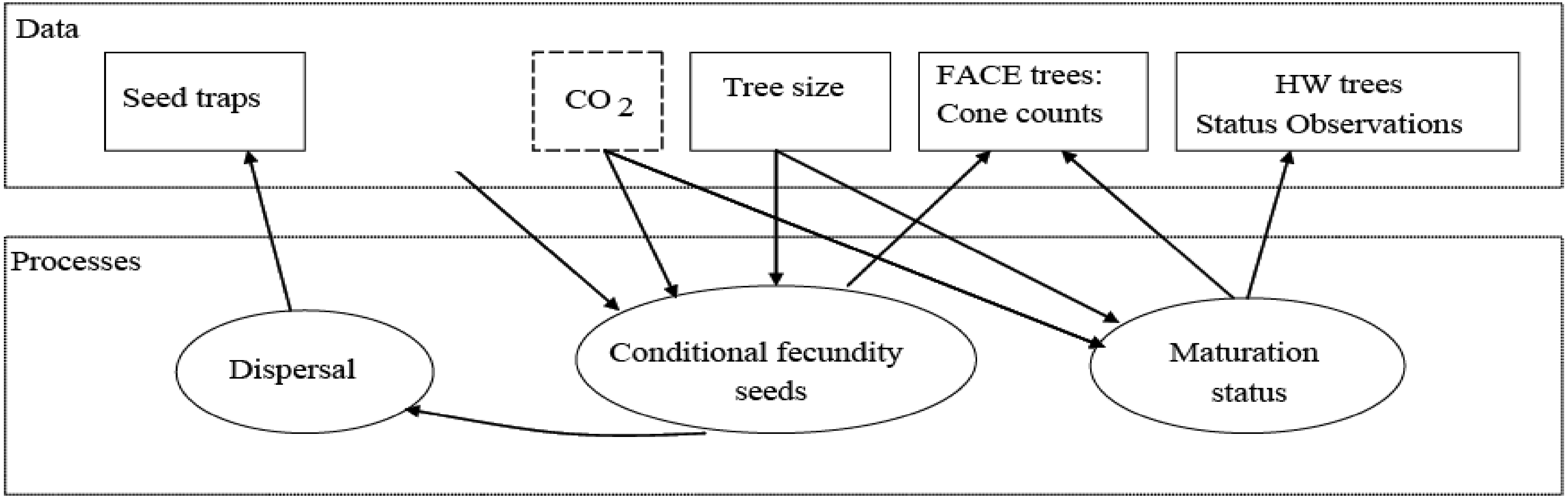
Fecundity of trees is often related to **tree size**:
seed #= fxn(**diameter** + treatment)



Variability Among Individuals – not random noise



Modeling Fecundity



- Cones and seeds inform (latent) fecundity estimate

- There are different ‘reasons’ for recording 0 cones/seeds…
    - Weather events
    - (latent) maturation status