ADVANCED BAYESIAN MODELING

Beyond Linear Regression

Linear Regression

$$X_i \;=\; (x_{i1}, \ldots, x_{ik}), \quad y_i \qquad\qquad i = 1, \ldots, n$$

X has k explanatory variables (usually including an intercept)

y is the response

In *linear* regression,

$$\mathrm{E}(y_i \mid \theta, X_i) \;=\; \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

Necessary for linear regression:

- y is numerical.
- y values cannot be too near any natural boundaries (e.g., 0).

  (Linear function $\beta_1 x_{i1} + \cdots + \beta_k x_{ik}$ could go outside the boundaries.)

What if these conditions aren't met?

Examples

- Cancer drug trial

  Response: Survived or not

  Not numerical (but could be converted to $0/1$)

- Accidents at an intersection

  Response: Number of accidents each day

  Numerical, but often at lower bound of 0

Note: Simply transforming the response does not help.

Data Example: Ebola Outbreaks

$$\begin{aligned} y_i &= \text{deaths in outbreak } i\text{, out of } n_i \text{ cases} \\[2em] x_{i1},\, x_{i2},\, x_{i3} &= \text{indicators of virus type (BDBV, EBOV, SUDV)} \\ x_{i4} &= \text{year outbreak began} \end{aligned}$$

Does fatality rate vary by virus type? Over time?

$$y_i \sim \mathrm{Bin}(n_i,\ ?\ )$$

Data Example: Ship Damage

$$y_{ijk} \quad = \quad \text{number of damage incidents to ships in category } i, j, k$$

i : type of ship $\qquad$ j : era of construction $\qquad$ k : period of operation

How does rate of damage incidents depend on these factors?

$$y_{ijk} \quad \sim \quad \text{Poisson(?)}$$

Want a more general type of regression that:

- Preserves linear dependence on explanatory variables
- Naturally allows response distributions other than normal

Motivates *generalized linear models* ...