Expanding the linear regression framework

- systolic blood pressure
- beetle death
- crab satellites
- time (blood coagul)
- defoliation of trees
- reaction time
- species on beaches

Y ← the response, dependent variable

random variable

- continuous
- discrete : binary, count
- categorical : ordered / unordered

our main modelling objective

AIMS:

1) Make a prediction of Y in a specific setting.

2) Understand relationship between response and

- ori
- area, location
- dose
- day
- smoke status
- where on beach
- days without sleep

$X_1, X_2, \ldots, X_k$

- continuous
- discrete
- categorical : nominal / ordered
- factor

covariates, explanatory variables independent variables, regressors, predictors, ...

fixed – not random. If random, we consider conditional modelling

We imagine collecting data and let i denote obs i.

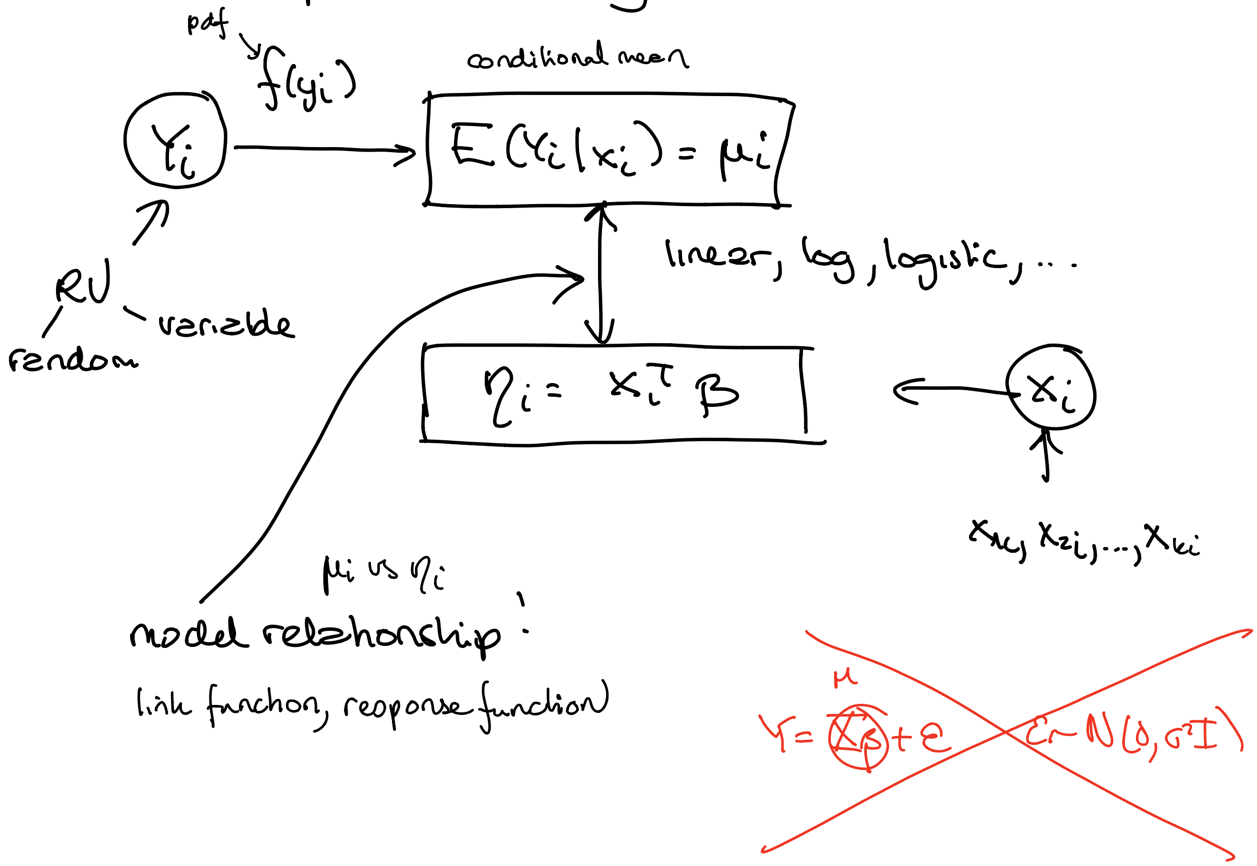
We will consider relationships between

the conditional mean of Y_i : $E(Y_i|x_i) = \mu_i$

and linear combinations of the covariates in a linear predictor:

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} = \underset{1\times p}{x_i^T}\,\underset{p\times 1}{\beta}, \qquad p = k+1$$

Pairs (x_i, Y_i) are in most cases independent $i = 1, \dots, n$ (so (x_1, Y_1) independent of (x_2, Y_2)) but also dependencies may be modelled.



Univariate exponential family ← Y_i response variable

(pdf, pmf)

$$f(y_i \mid \theta_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} \cdot w_i + c(y_i, \phi w_i) \right)$$

θ_i canonical parameter : parameter of interest [N: μ_i]

ϕ : nuisance (dispersion param) [N: σ^2]

[N: 1]

w_i : weight

b & c known functions.

Can be shown: $E(Y_i) = b'(\theta_i)$ and $Var(Y_i) = b''(\theta_i) \cdot \frac{\phi}{w}$

tomorrow

$\frac{db}{d\theta_i}$

Our models and results (^{M}2-5) made for such Y_i's.

Work more on this in IL