Some final thoughts

Patrick Breheny

April 25

Uses

- Linear and generalized linear models are vitally important to statistics and found in virtually all fields
- They are essential in observational studies to make adjustments for potentially confounding variables
- They are even useful in randomized trials to obtain more accurate predictions and efficient estimates
- Furthermore, the ideas of setting up a regression model (additive effects, interactions, indicators for categories, link functions, systematic/random components) appear in many other sorts of models, as we have seen with negative binomial/quasilikelihood/multi-category models and as you will see again next semester with survival models

Limitations

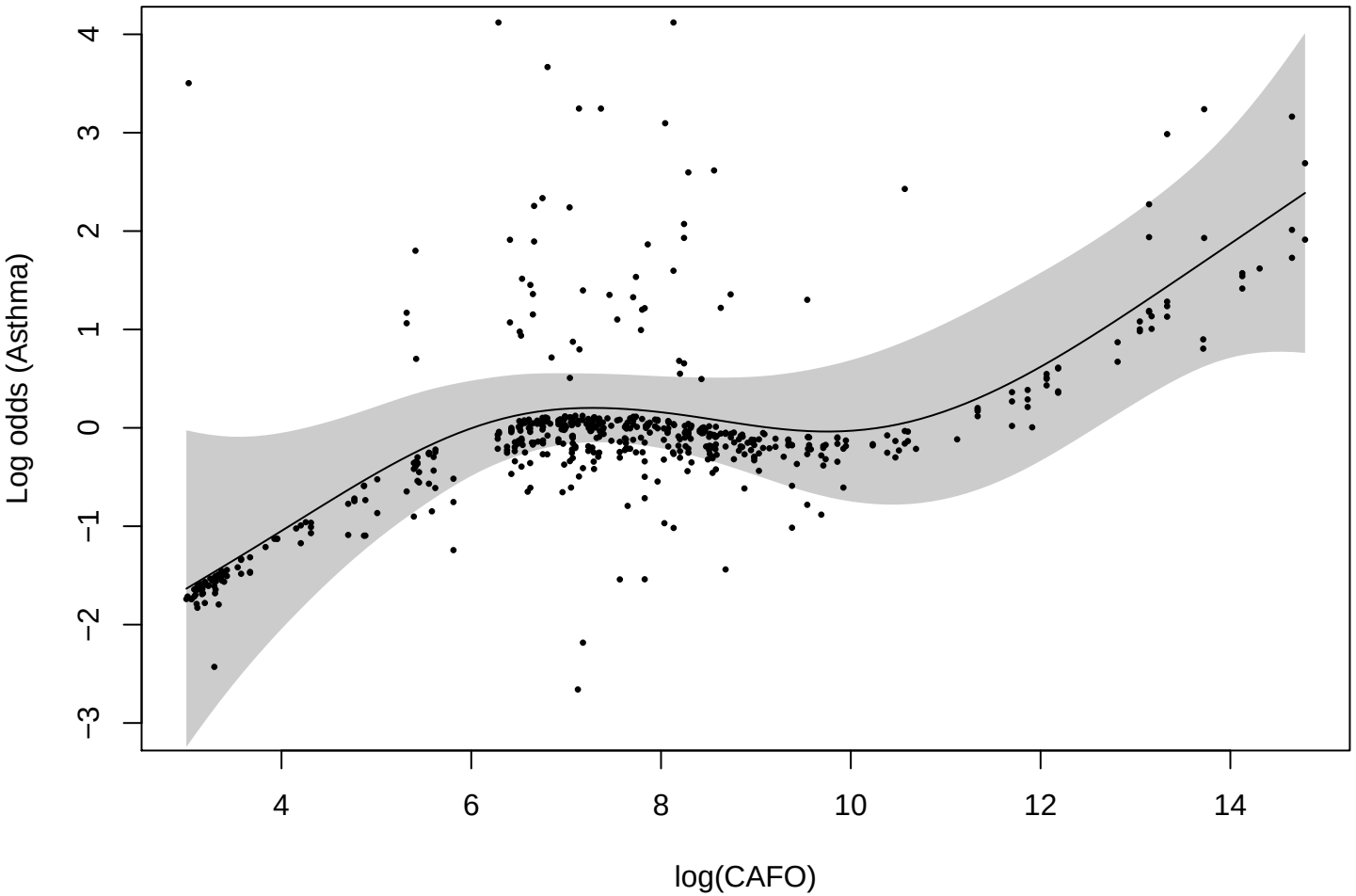
- At the same time, it is always important to keep in mind the limitations of modeling and the rather strong assumptions we often make in constructing a model, chief among them:
  - Linearity
  - Additivity
  - Lack of unobserved confounders
- Certainly, these are almost always violated in an absolute sense, but keep in mind:
  - The bias-variance tradeoff
  - "All models are wrong; some models are useful"

Non-linear models

- Linear and generalized linear models are certainly the most important class of models, but they are not the only kind of model
- For example, we sometimes wish to allow the effect of an explanatory variable to be a smooth curve rather than a line:

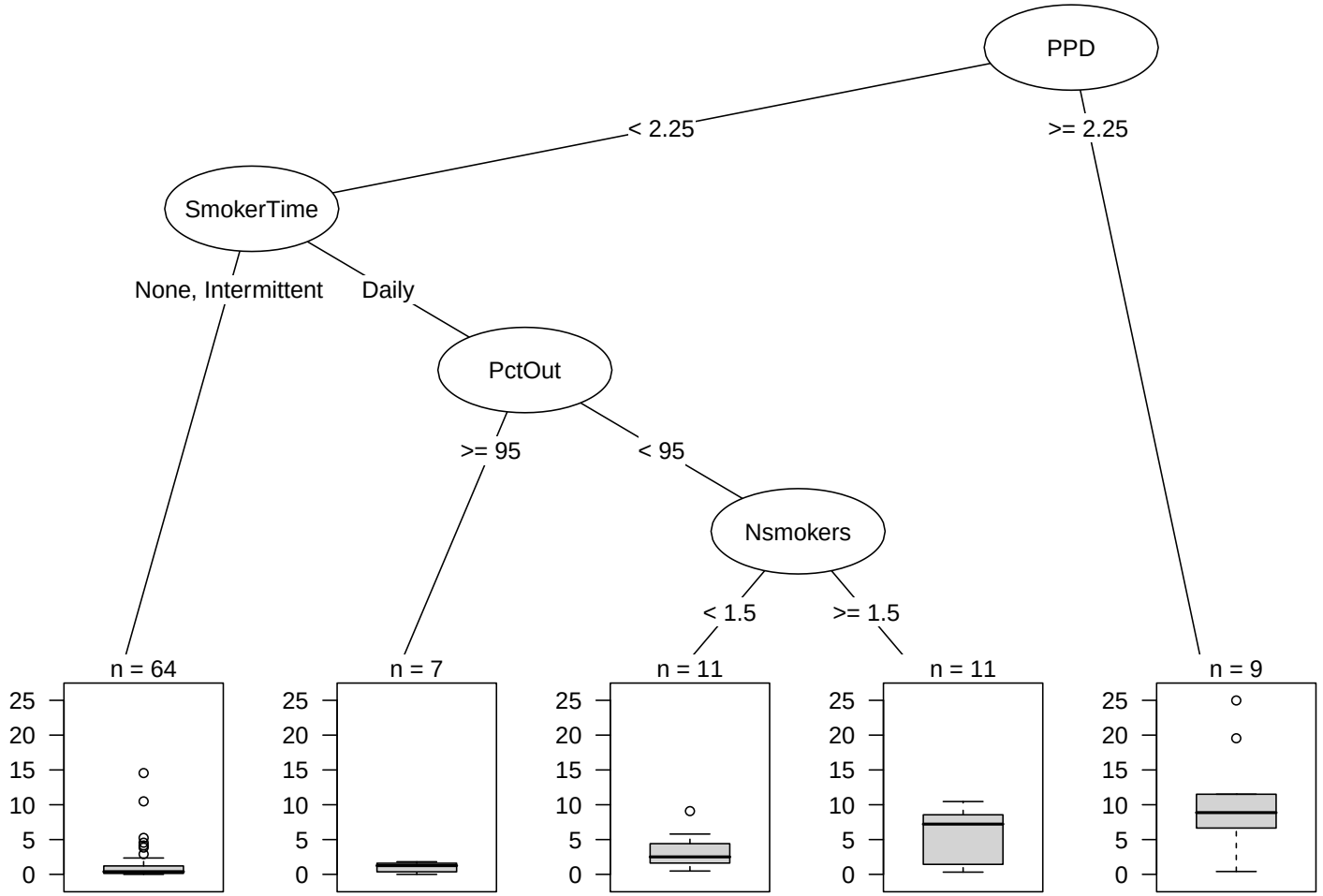
$$g(\mu_i) = f(x_i)$$

Non-linear models: Example



Tree-based models

There are also tree-based models:



And much more...

- And there are many additional extensions/modifications/alternatives that have been proposed as well:
  - Robust regression
  - Methods for dealing with highly correlated explanatory variables
  - Methods for variable selection that account for uncertainty in the model selection process
  - ...