

Introduction to beta regression in glmmTMB

Mollie Brooks

June 2023

This video covers

- basics of the beta distribution
- an example of gasoline yield
 - with a model on the dispersion parameter
- an example of percent cover of seagrass beds in the Northeast USA
 - zero-inflation
 - residual tests via DHARMA
 - splines

Basics

- `beta_family(link = "logit")`
- good for proportion data (between 0 and 1)
- parameterization like `betareg` and Ferrari and Cribari-Neto (2004)
- conditional variance $\frac{\mu(1-\mu)}{1+\phi}$
- s.t., increasing ϕ decreases the variance

Ex: Gasoline (Prater 1956)

- from the betareg package
- proportion of crude oil after distillation and fractionation
- includes a model on the dispersion parameter
- temperature affects yield and variability

Prater, N.H. (1956). Estimate Gasoline Yields from Crudes. *Petroleum Refiner*, 35(5), 236–238.

see `code_beta.R`

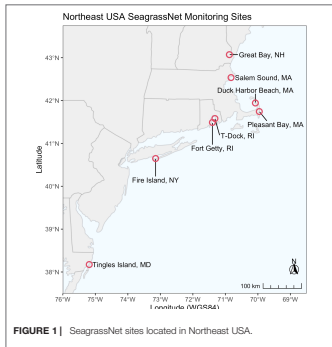
Beta does not allow 0s and 1s

- proportion data (between 0 and 1)
- if not too many 0s & 1s then remove them using:
 - `n = nrow(dat)`
 - `newprop = prop*(n-1)+1/2)/n`

Smithson and Verkuilen 2006 “A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables”

Ex: Eelgrass affected by temperature (Plaisted et al. 2022)

- eelgrass, *Zostera marina* L.
- 8 location with 12 quadrats per location monitored annually
- percent cover, including 0s and 100s
- affected by summer water temperature one year prior



see `code_betazi.R`

Recap

- dispersion modeled with a covariate
- quadrats nested in locations
- eelgrass ex: ZI better than “squeezing” 0s
- predict from the conditional and ZI models separately
- `splines` package allows flexible nonlinear functions
- `DHARMa` package for checking residuals