NRES_798_11_201501

Generalized linear model

Logistic regression

The General Linear Model

In a general linear model

$$= \quad _0 + \quad _1 \quad + \quad + \qquad +$$

the response y_i is modelled by a linear function of explanatory variables x_i, plus an error term

General and Linear Model

Here general refers to the dependence on potentially more than one explanatory variables, v.s. the simple linear model:

$$= {}_0 + {}_1 +$$

The model is linear in the coefficients,

but not

Error structure

We assume that the errors ϵ are independent and identically distributed after

$$E[\epsilon_i] = 0$$

and $\text{var}[\epsilon_i] = \sigma^2$

Typically we assume

$\epsilon_i \sim N(0, \sigma^2)$

on parameters

a t-test

Restrictions of Linear Models

Although a useful framework, there are some situations where general linear models are not appropriate

- the range of Y is restricted (e.g. binary, count)
- the variance of Y depends on the mean

Generalized linear models extend the general linear model framework to address both of these issues

GLM potential response variables

Count data expressed as proportions

Generalized Linear Models (GLMs)

A generalized linear model is made up of three things:

a linear predictor

$$= \quad _0 + \quad _1 \quad + \quad +$$

and two functions

a link function that describes how the mean, E(Yi) = ui depends on the linear predictor

$$(\quad) =$$

An variance function that describes how the variance, var(Yi) depends on the mean

$$(\quad) = \quad (\quad)$$

where the dispersion parameter φ is a constant

Error structure

Normal General Linear Model as a Special Case

Error structure

Possible GLM error distributions

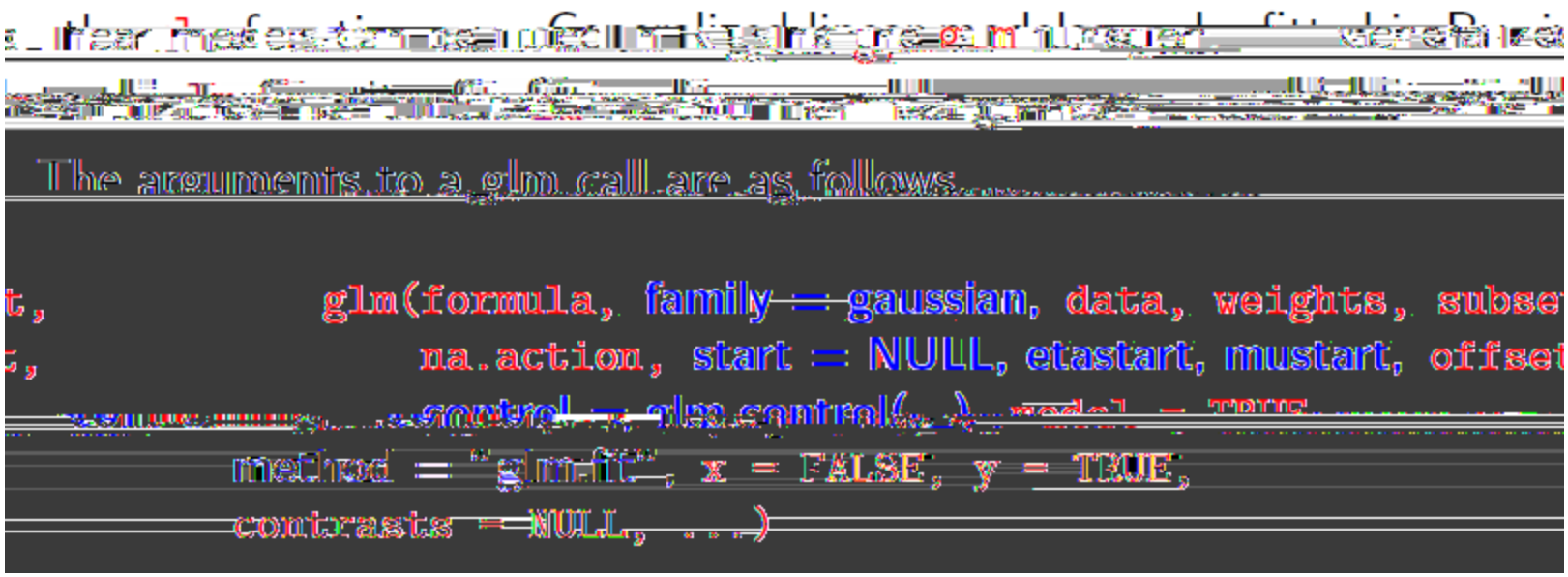
Transformations vs. GLM

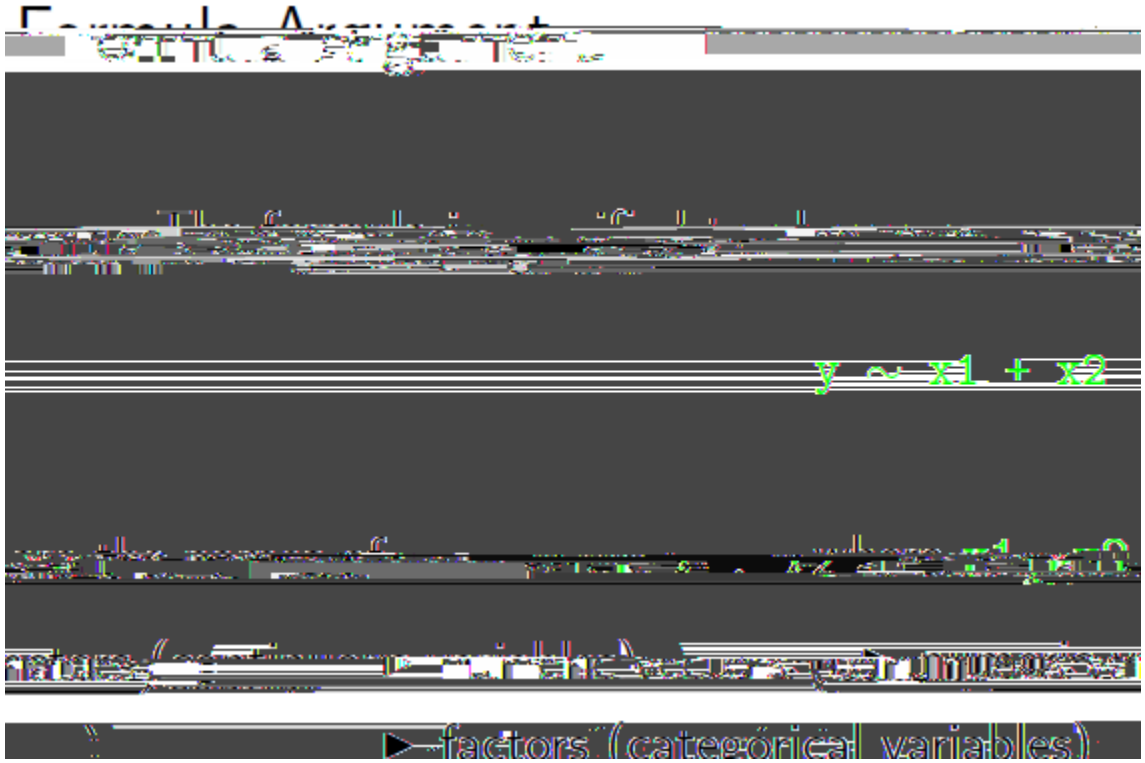
In some situations a response variable can be transformed to improve linearity and homogeneity of variance so that a general linear model can be applied.

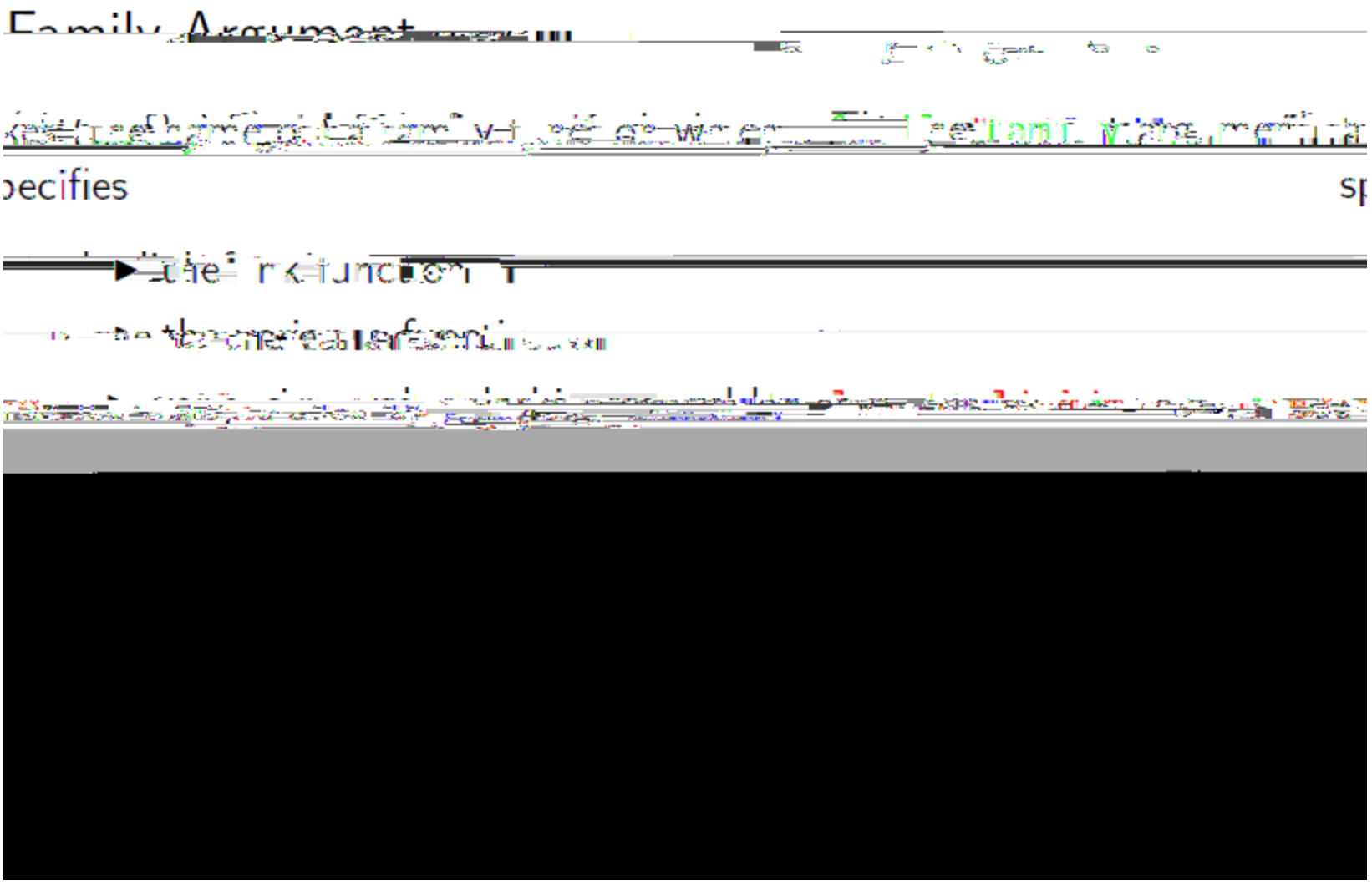
This has some drawbacks

- response variable has changed!
- transformation must simultaneously improve linearity and homogeneity of variance
- transformation may not be defined on the boundaries of the sample space

The glm Function

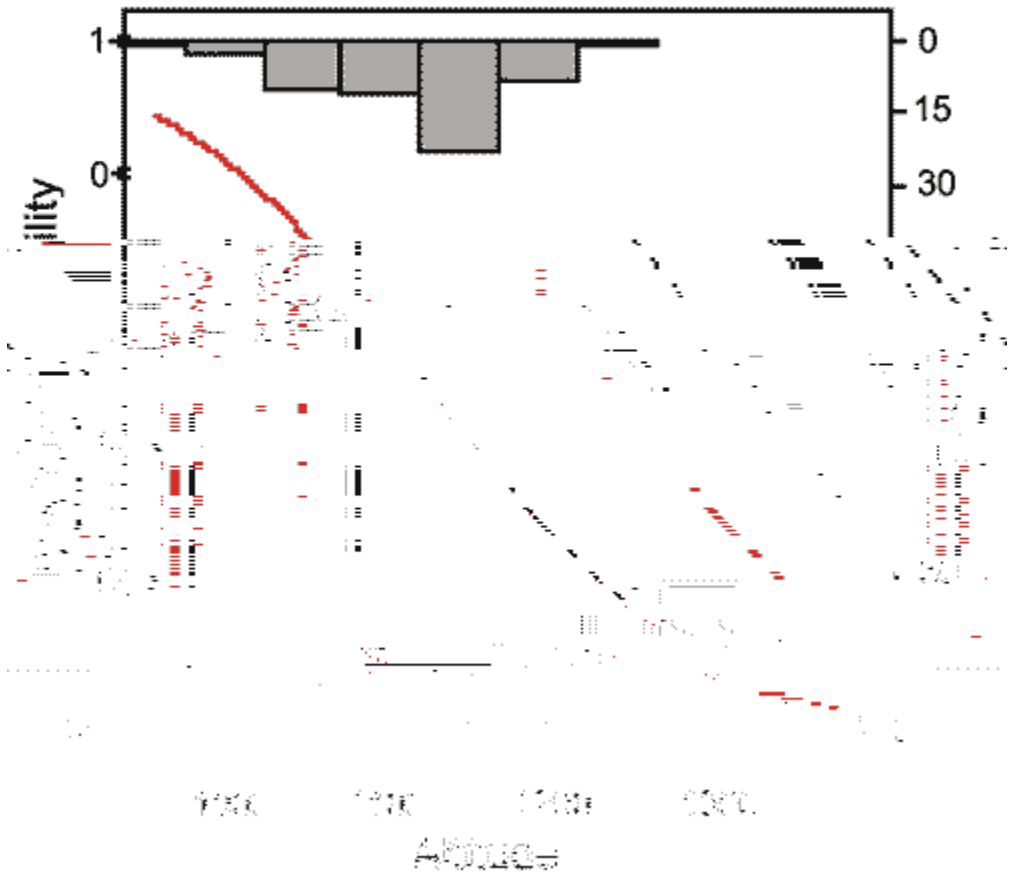



y ~ x1 + x2
factors (categorical variables)

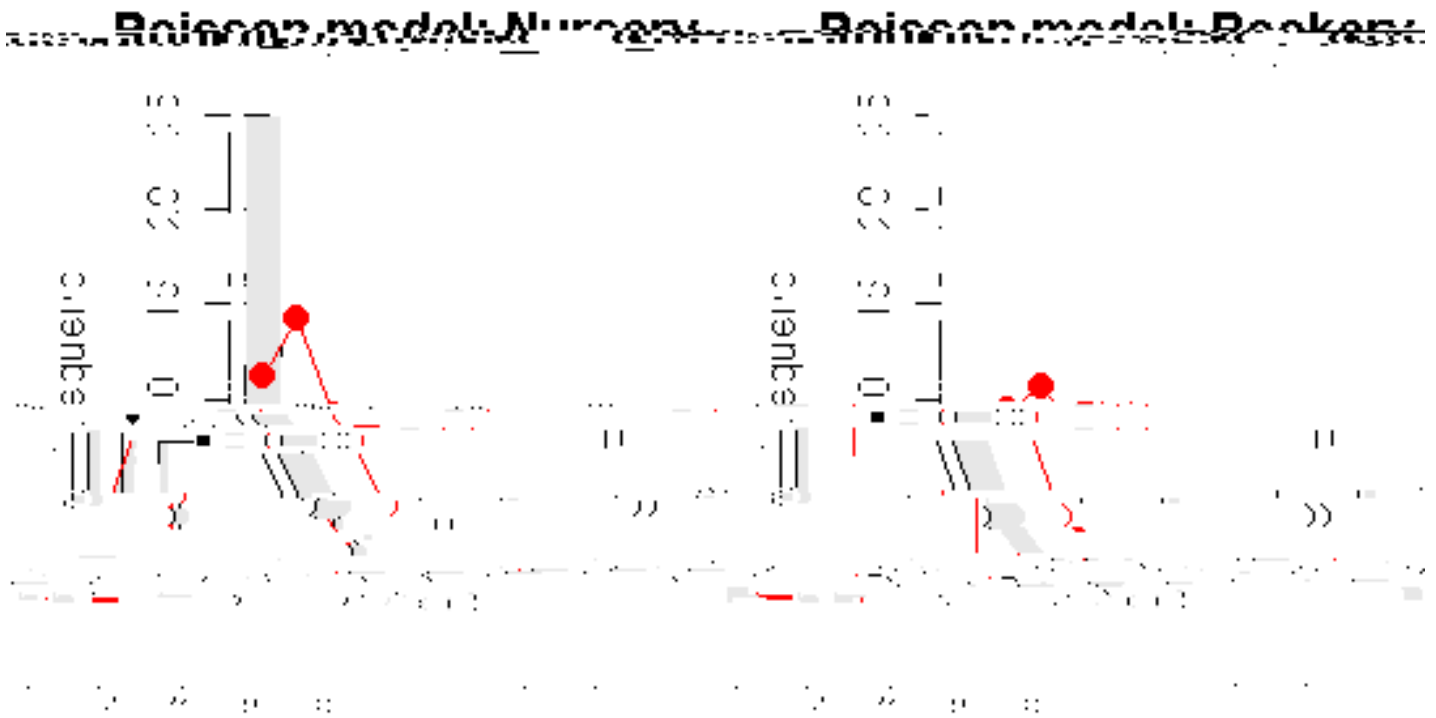


GLM: Logistic regression, binomial family

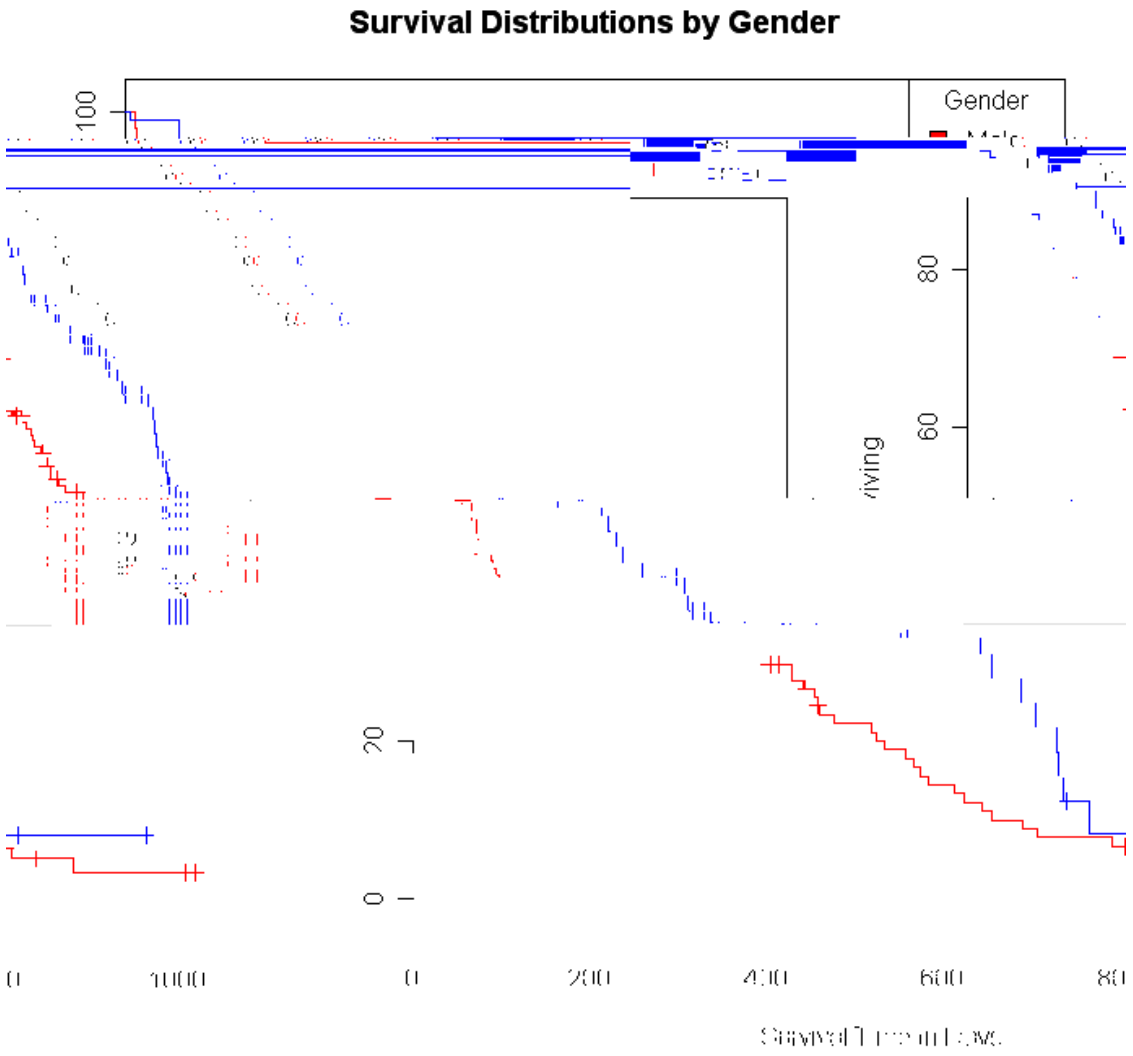
Probability of Norway spruce occurrence along an altitudinal gradient



GLM: Poisson regression, Poisson family



GLM: Survival analysis



Exponential distribution
Weibull distribution
Gamma distribution