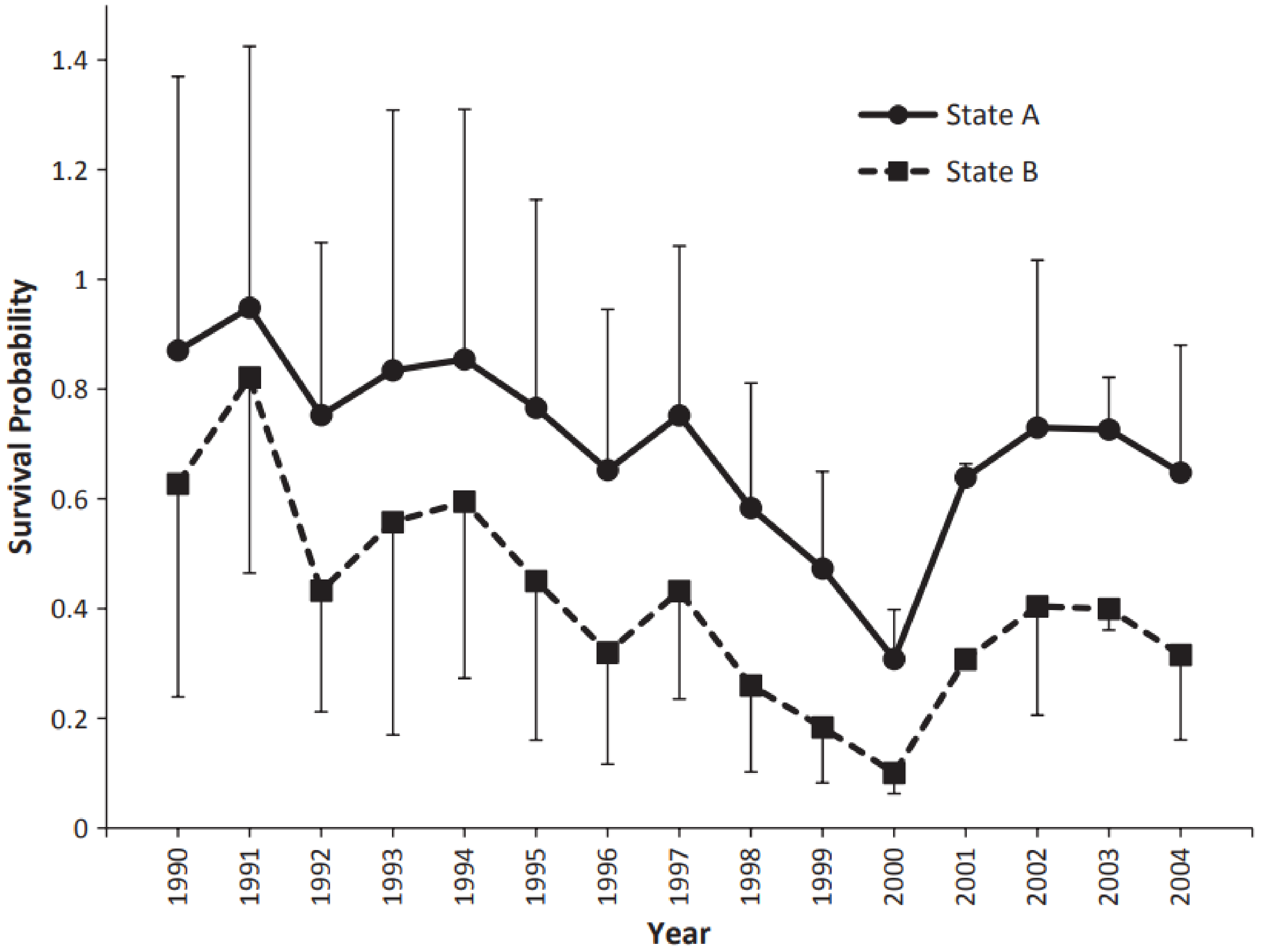
Link functions: constraining parameter space to the possible



But first, let's talk about 'assignments'

Week 3	Sep 10 Sep 12	linear models: lm() vs. brms() vs. custom Random and fixed effects	Potential proposal topic due

Table 2. Grading.

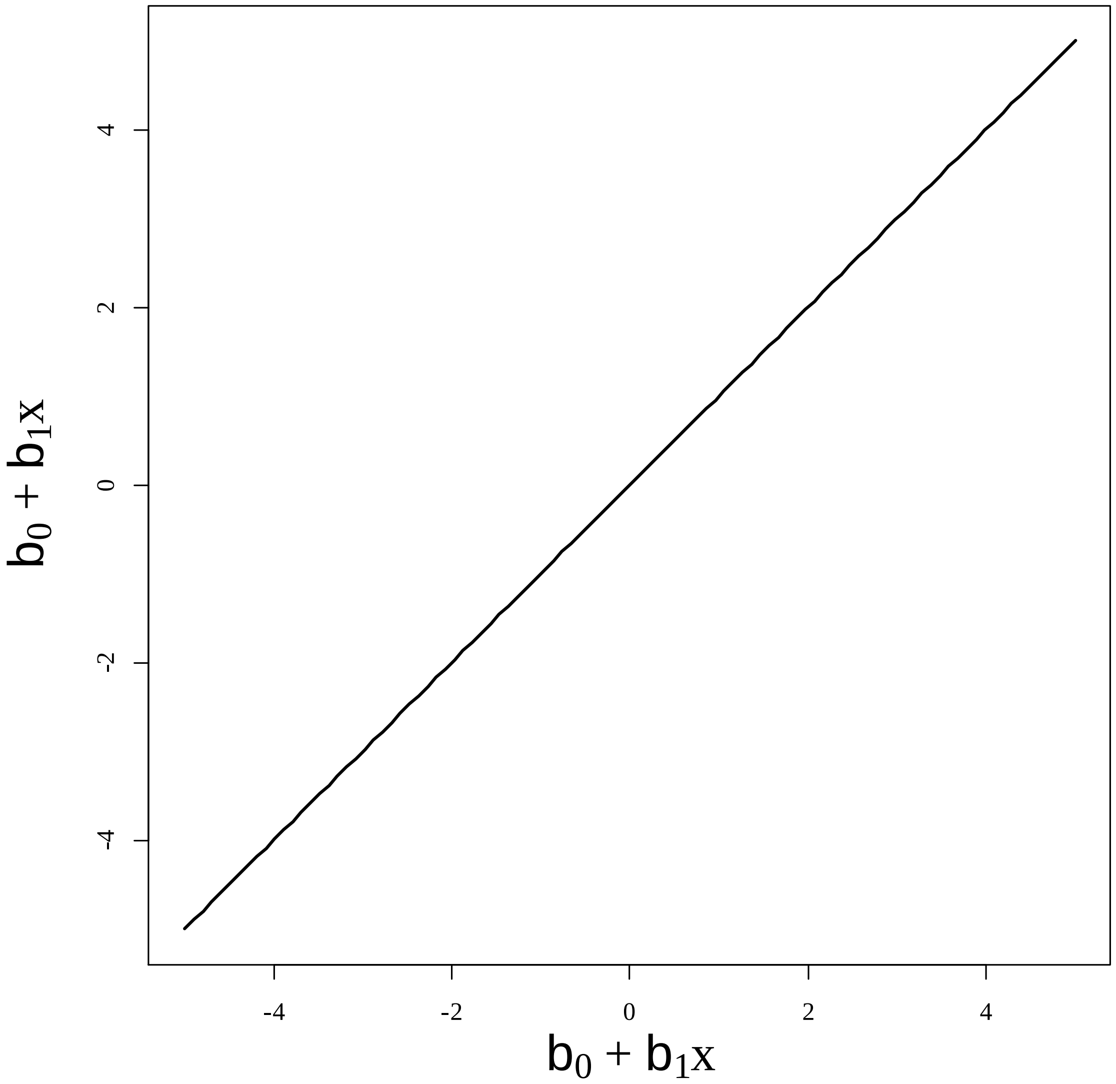
Category	Description	Date	Points	%
Project	Potential topic		10	5
	Final topic		10	5
	Final project		80	40
Homeworks			60	30
Participation	Contributing to discussion and labs		40	20
Total			200	

And a promise…

Thus far we’ve discussed linear models

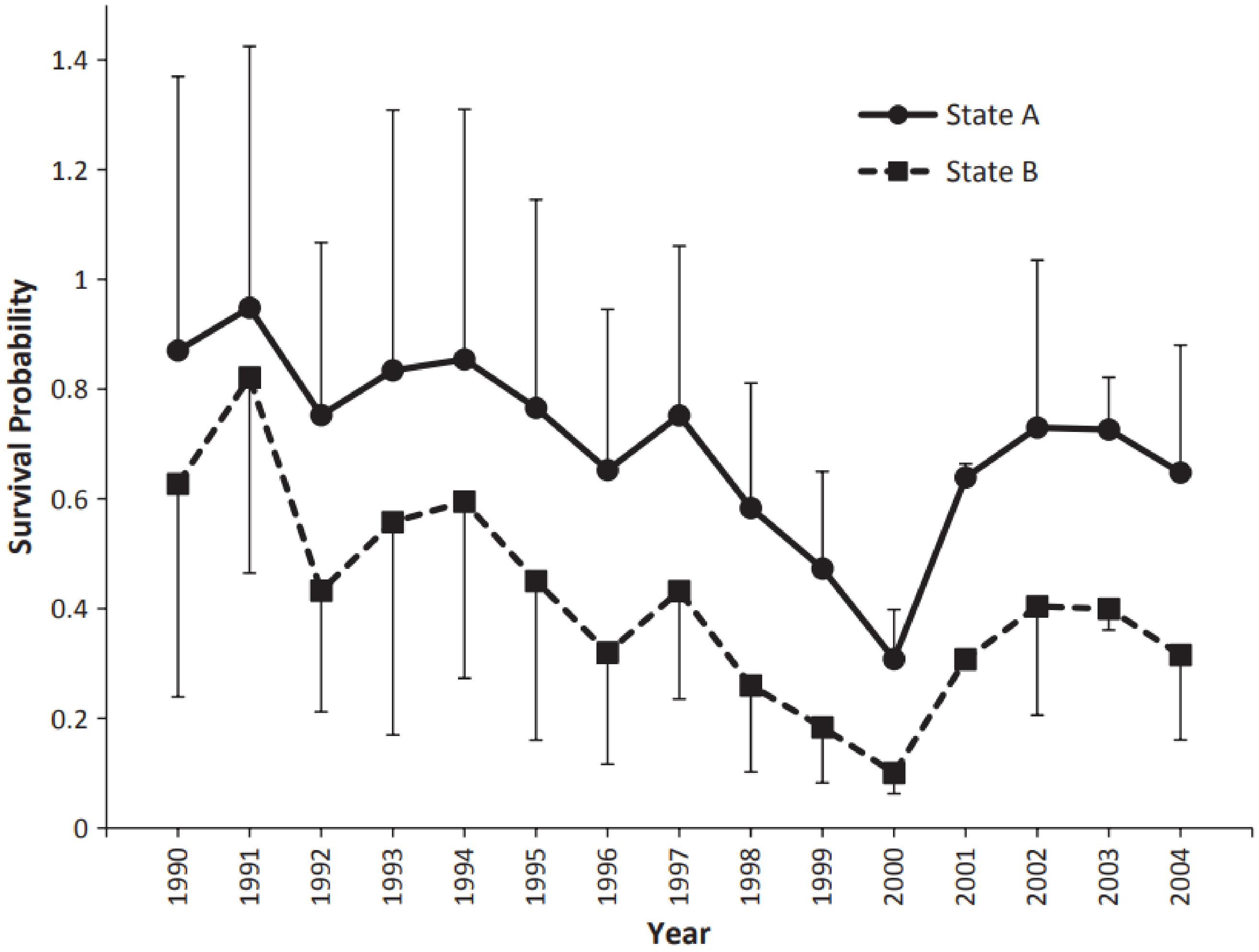
Linear models use an ‘identity link’

The scale of our response variable [-∞, ∞]



The scale of our model [-∞, ∞]

Transformations are often necessary [take a close look]



Lindberg [43, 4755], Sedinger [53; 7668], and Lebreton [76; 22357] **(2013)** *Ecology and Evolution*

Transformations are often necessary

How do we prevent ourselves from saying crazy [stuff]?

Let’s imagine some ecological data types

Let’s imagine some ecological data types

1. Counts of individuals [discrete, never less than 0?!]

2. Lengths, weights, or other morphometric measurements [continuous, never less than 0?]

3. Number of eggs or offspring [discrete, never less than 0]

Let’s imagine some ecological data types

1. Proportion or percent cover of something or other

2. Survival [yes/no; 1/0]

3. Pregnancy [yes/no; 1/0]

4. Hatching [yes/no; 1/0]

5. Morph [categorical?; 1/0 or 1, 2, ... , n]

All of these data types share some common attributes

1. You can’t have < 0’s
2. Many can only be 0’s or 1’s

We often need to constrain parameters to be > 0 or between 0 and 1.

So how do we handle that in our models?

Link functions!

How the heck do we do that?

Link functions: constraining parameter space to the possible

Or, what is a natural log?

What is e? [an irrational and transcendental number]

e is the basis of many of the link functions used in Ecology

Link functions allow us to <u>constrain</u> parameter estimates so we don't predict things like 150% survival, 200% shrub cover, or that an alligator that is -2 ft long will weigh -100 pounds

M. S. Lindberg *et al.*

Heterogeneity and Harvest Dynamics

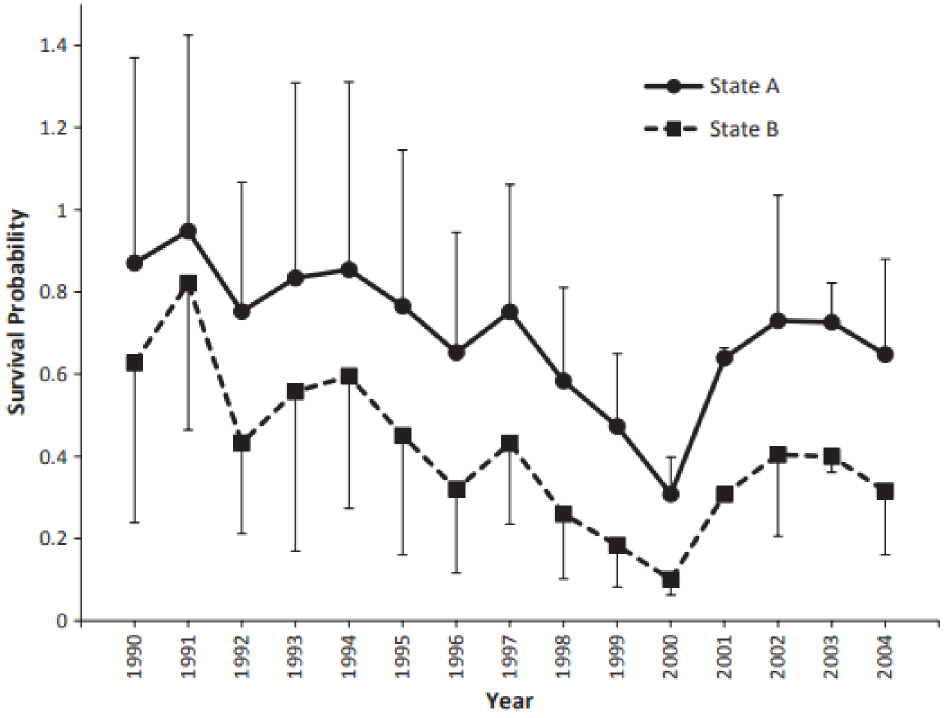


Figure 2. First-year survival probability of brant for states A and B. Errors bars are ±1 standard error in single direction for clarity.

What is e? [an irrational and transcendental number]

Imagine you have a bank account with $1 in it that earns 100% interest [put more money in the account!]

If you calculate interest annually, at the end of the year, you'll have $2

$\$1 \times 2^1 = \$2$

What is e? [an irrational and transcendental number]

Imagine you have a bank account with $1 in it that earns 100% interest [put more money in the account!]

If you calculate interest biannually, at the end of the year, you’ll have $2.25

$\$1 \times 1.5^2 = \$2.25$

What is e? [an irrational and transcendental number]

Imagine you have a bank account with $1 in it that earns 100% interest [put more money in the account!]

If you calculate interest quarterly, at the end of the year, you’ll have $2.44

$\$1 \times 1.25^4 = \$2.44$

What is e? [an irrational and transcendental number]

Compounding annually yields 2
Compounding biannually yields 2.25
Compounding quarterly yields 2.4414…
Compounding monthly yields 2.613035….
Compounding weekly yields 2.692597…
Compounding daily yields 2.714567…

Notice a pattern?

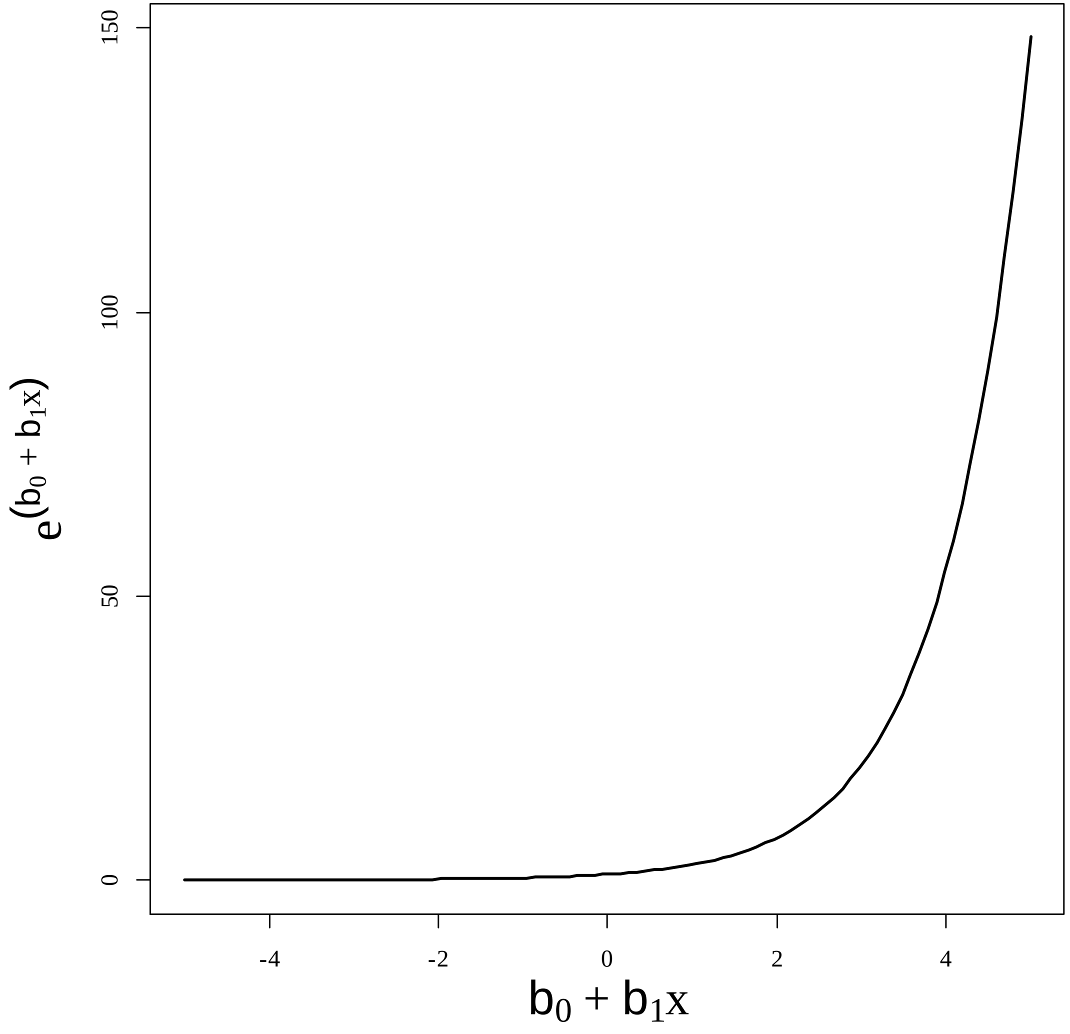
$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n}$$

Compounding **continuously** yields e

2.718281828459045235360287471352662497757247093699959574966967627724
0766303535475945713821785251664274274663919320030599218174135966290
4357290033429526059563073813232862794349076323382988075319525101901
1573834187930702154089149934884167509244761460668082264800168477411
8537423454424371075390777449920695517027618386062613313845830007520
4493382656029760673711320070932870912744374704723069697720931014169
2836819025515108657463772111252389784425056953696770785449969967946
8644549059879316368892300987931277361782154249992295763514822082698
9519366803318252886939849646510582093923982948879332036250944311730
1238197068416140397019837679320683282376464804295311802328782509819
4558153017567173613320698112509961818815930416903515988885193458072
7386673858942287922849989208680582574927961048419844436346324496848
7560233624827041978623209002160990235304369941849146314093431738143
640546253152096183690888707016768396424378140592714563549061303107

We can do beautiful things with *e* (constrain to > 0)

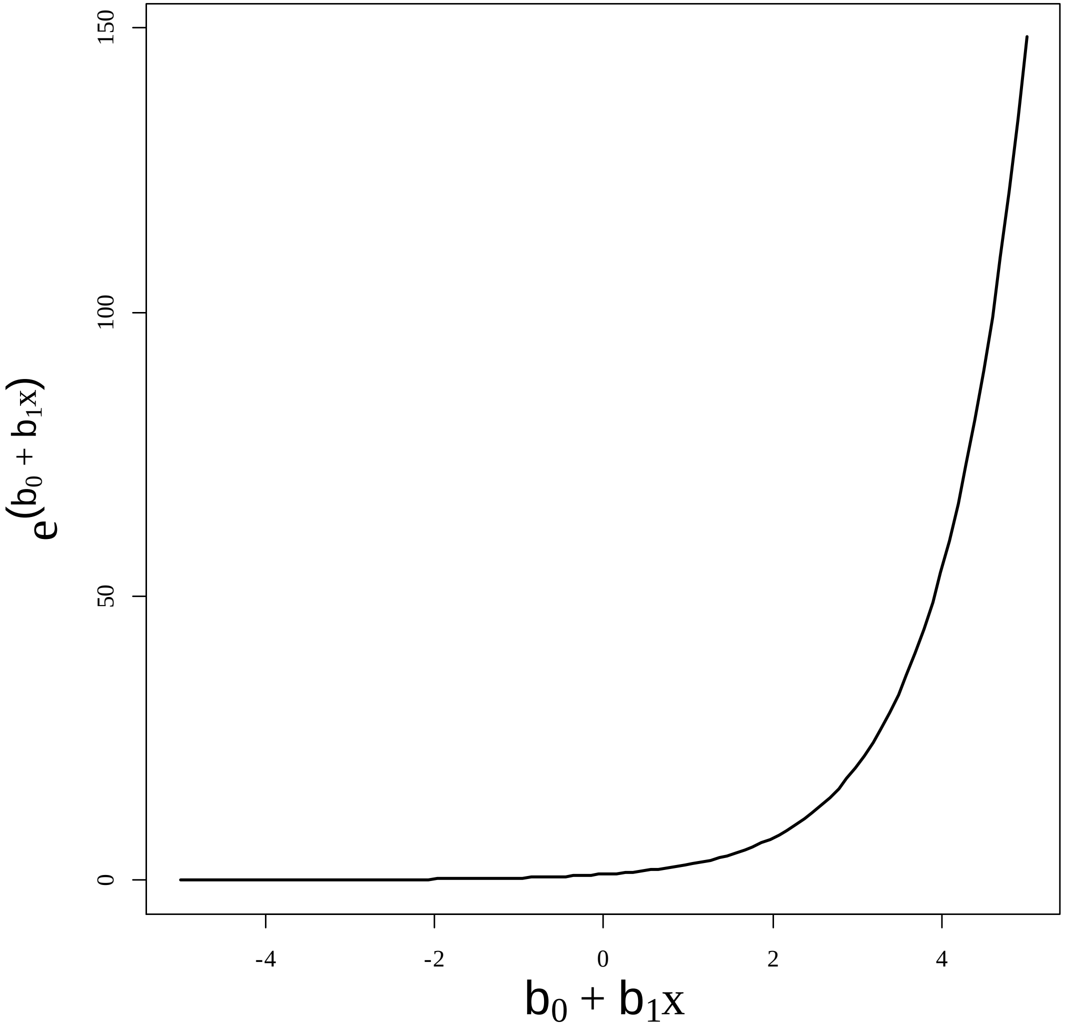
The scale of our response variable [0, ∞]



The scale of our model [-∞, ∞]

We can do beautiful things with *e* (constrain to > 0 & < 1)

The scale of our response variable [0, ∞]

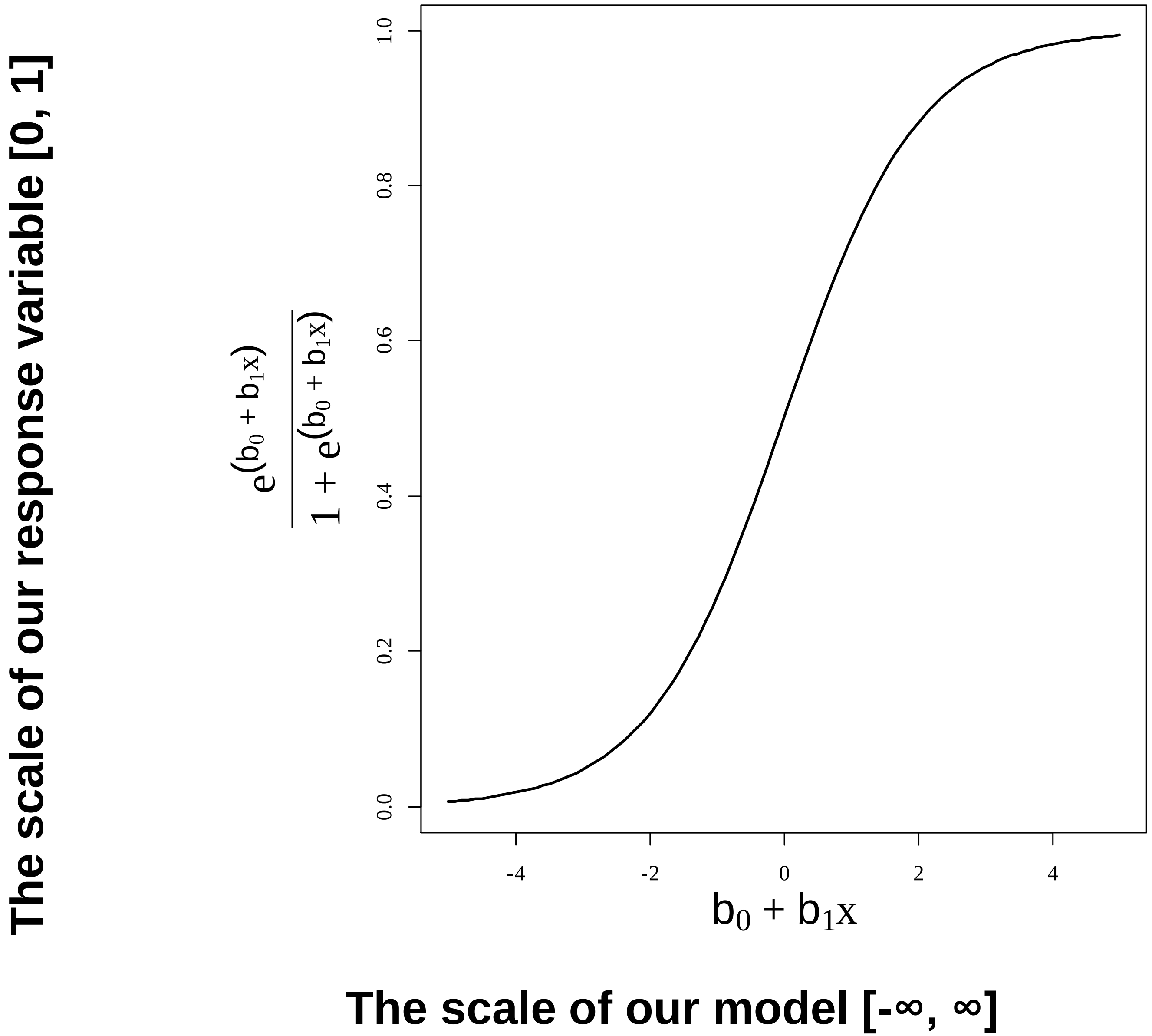


The scale of our model [-∞, ∞]

$$y = e^{\beta_0 + \beta_1 \times x}$$

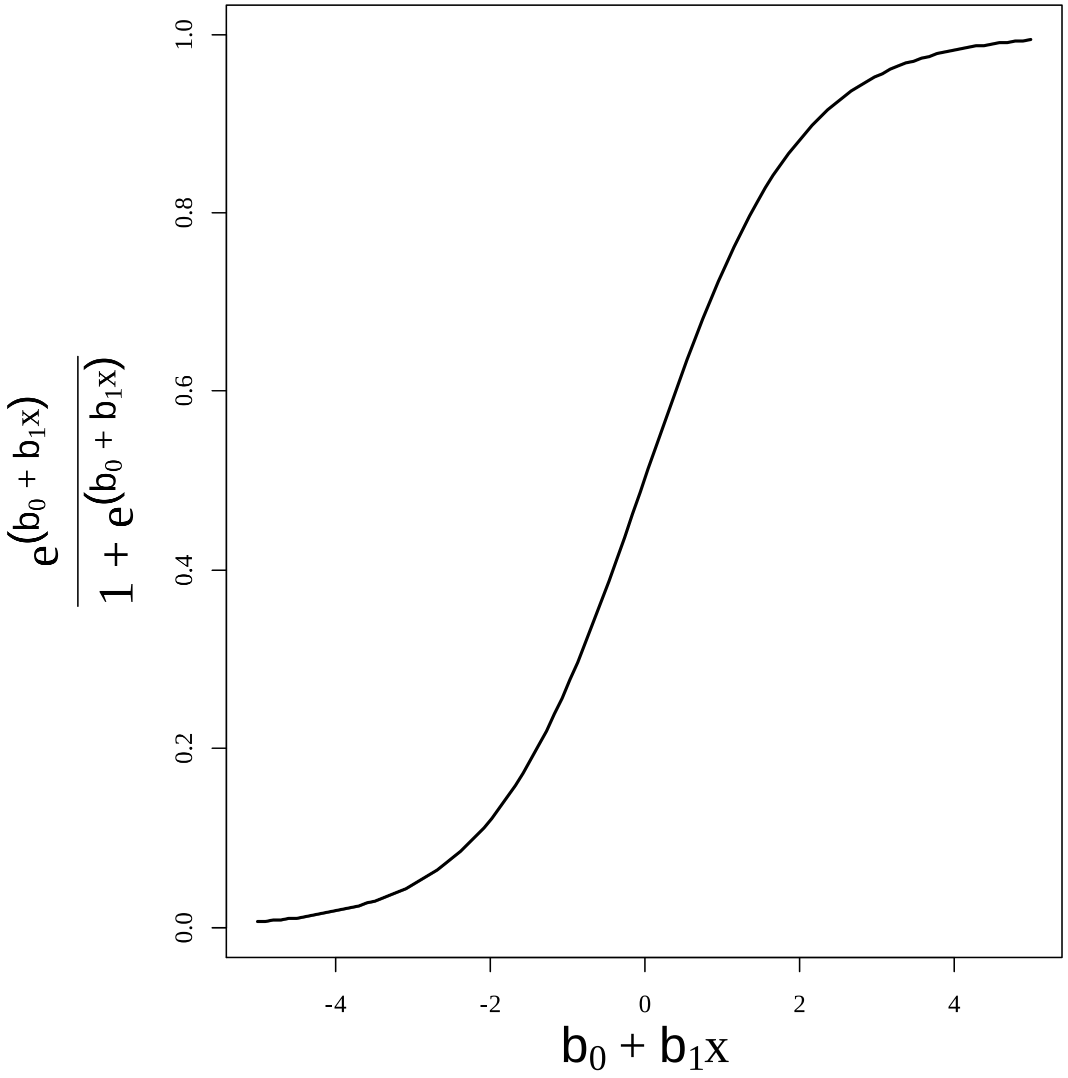
$$\beta_0 + \beta_1 \times x = \ln(y)$$

We can do beautiful things with *e* (constrain to > 0 & < 1)



We can do beautiful things with *e* (constrain to > 0 & < 1)

The scale of our response variable [0, 1]



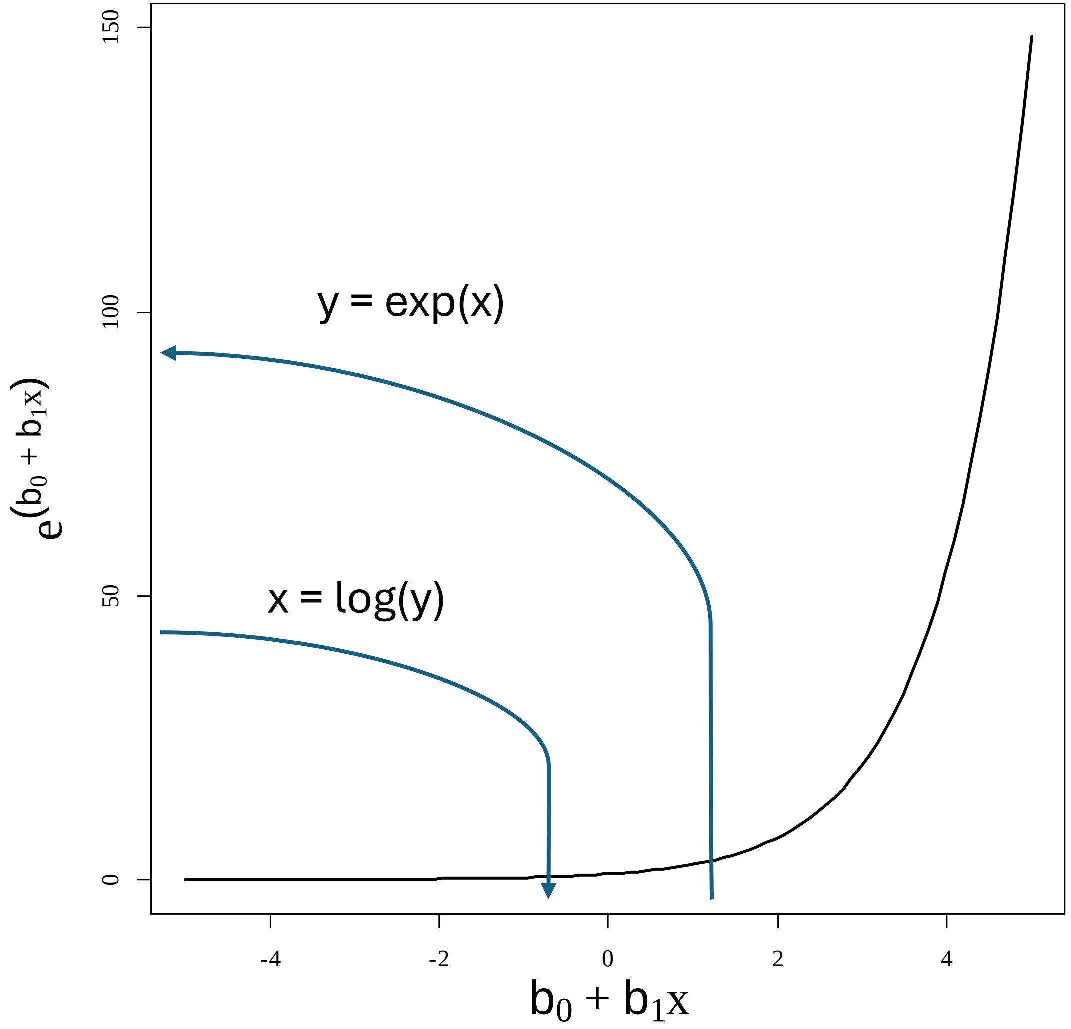
The scale of our model [-∞, ∞]

$$y = \frac{e^{\beta_0 + \beta_1 \times x}}{1 + e^{\beta_0 + \beta_1 \times x}}$$

$$\beta_0 + \beta_1 \times x = \ln\left(\frac{y}{1 - y}\right)$$

Implementing in base R (log link)

The scale of our response variable [0, ∞]

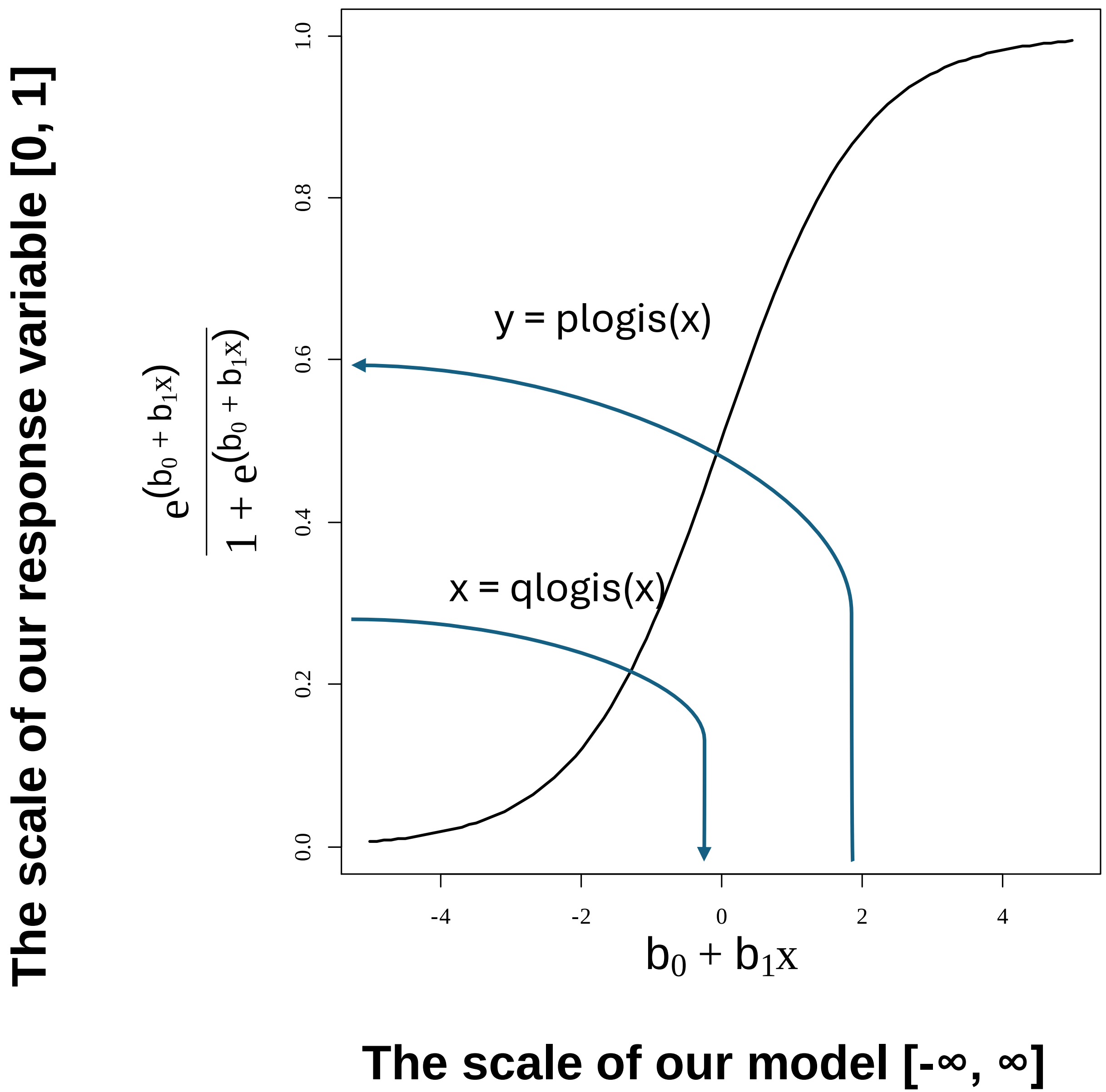


The scale of our model [-∞, ∞]

$$y = e^{\beta_0 + \beta_1 \times x}$$

$$\beta_0 + \beta_1 \times x = \ln(y)$$

Implementing in base R (logit link)



$$y = \frac{e^{\beta_0 + \beta_1 \times x}}{1 + e^{\beta_0 + \beta_1 \times x}}$$

$$\beta_0 + \beta_1 \times x = \ln\left(\frac{y}{1 - y}\right)$$

Today, we’ll reanalyze two datasets…

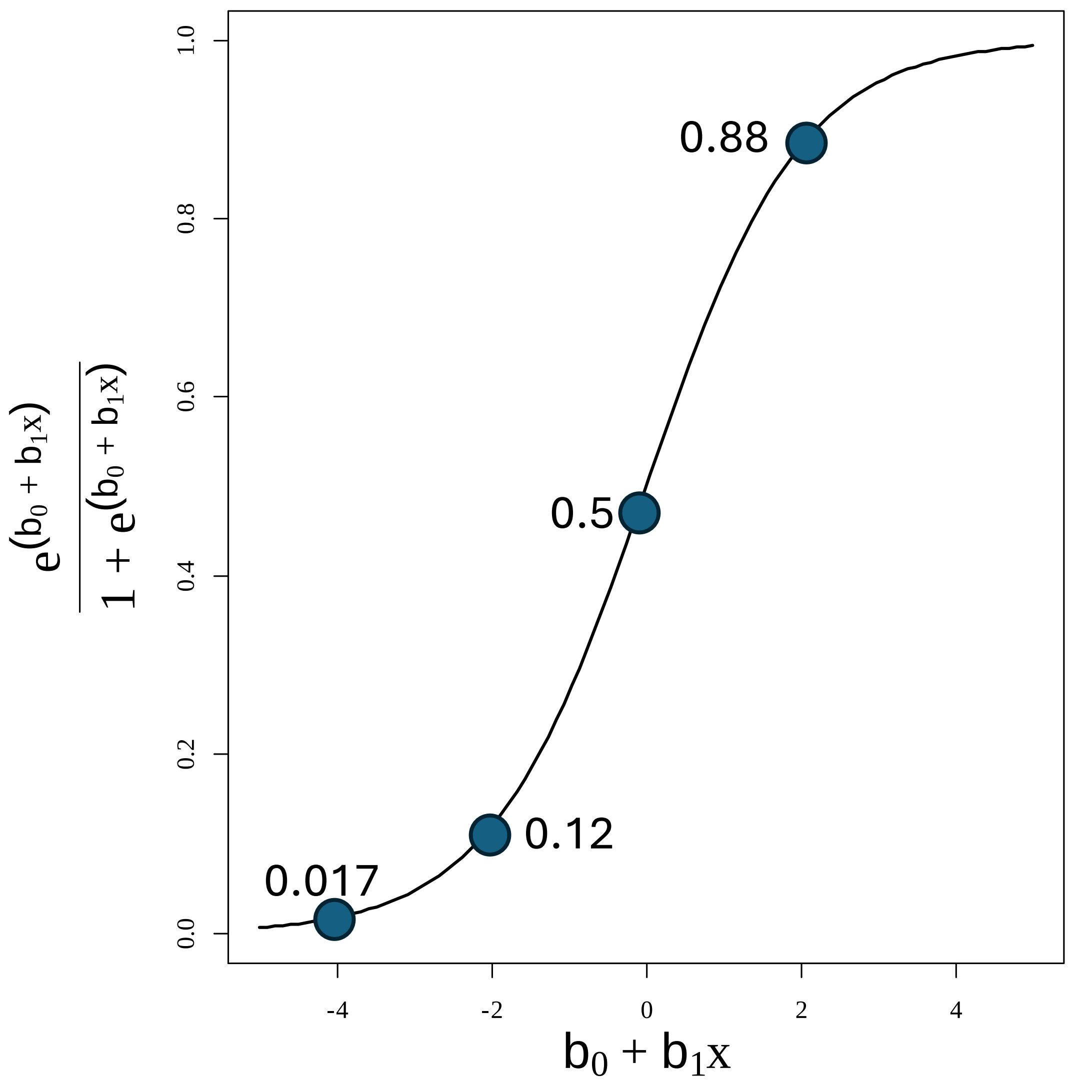
1. We’ll model the alligator mass ~ length data using a log-normal distribution (i.e., constraining mass to be positive!)

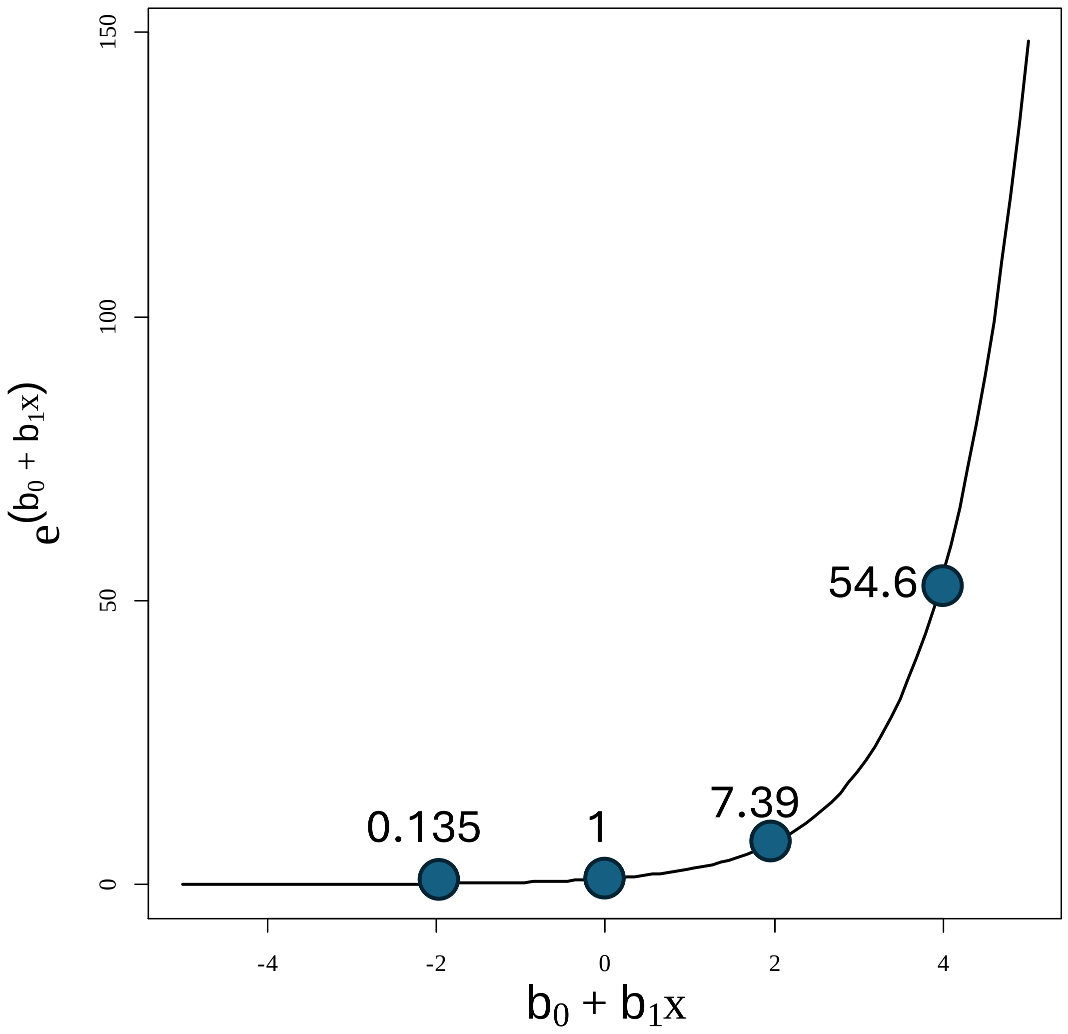
2. We’ll model some duck band-recovery data using a logit-link (i.e., constraining the probability of being shot, recovered, and reported to be > 0 and < 1)

Things to think about:

1. Beta estimates are no longer ‘cut and dry’ rise/run. They are relative to your position on the link function and the shape of the link function.

Beta estimates are no longer cut and dry!



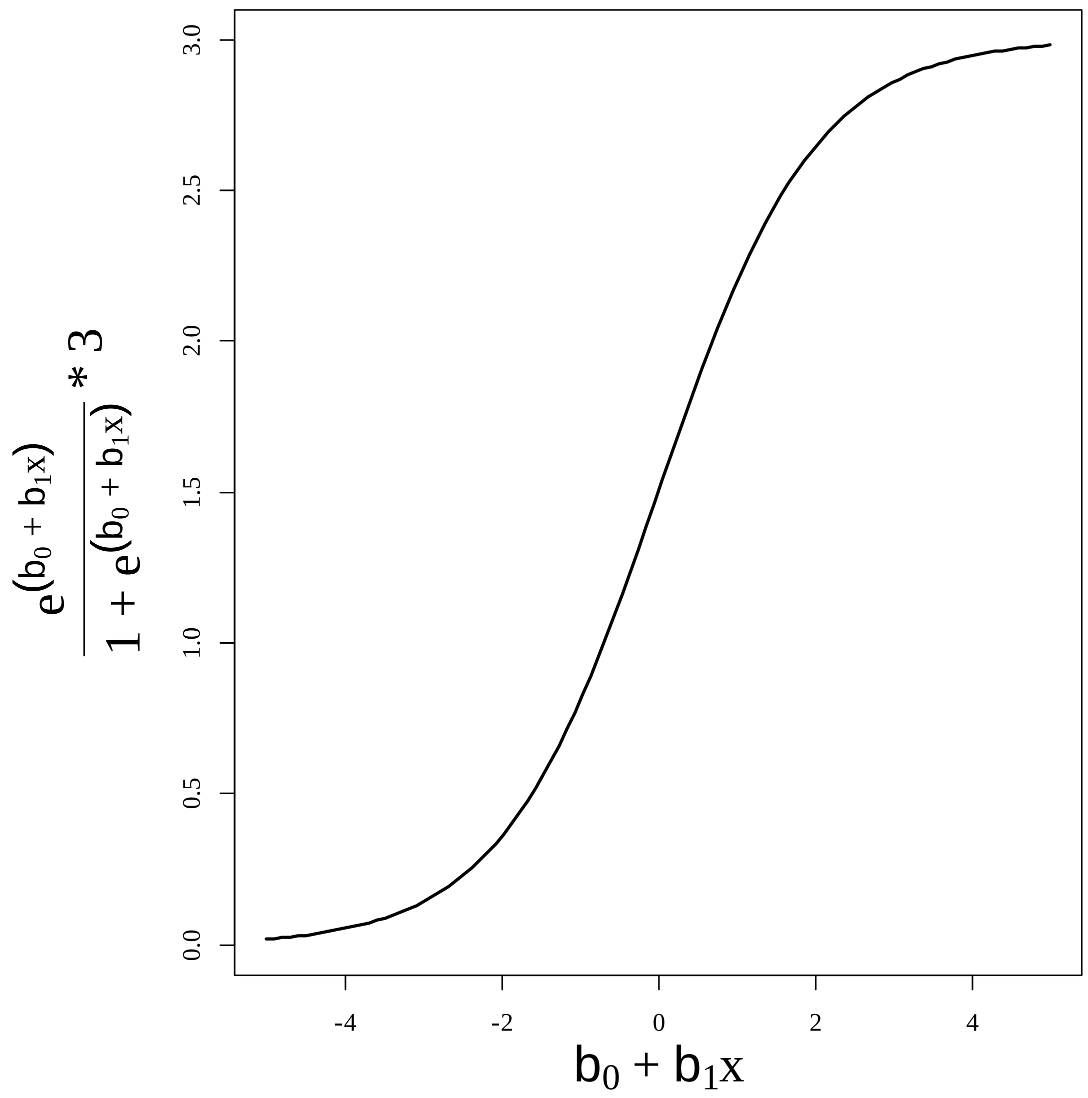


Things to think about:

1. Beta estimates are no longer ‘cut and dry’ rise/run. They are relative to your position on the link function and the shape of the link function.

2. There are many, many different types of link functions that do very similar things. Sometimes using a slightly different function can help with parameter estimation (related to point #1).

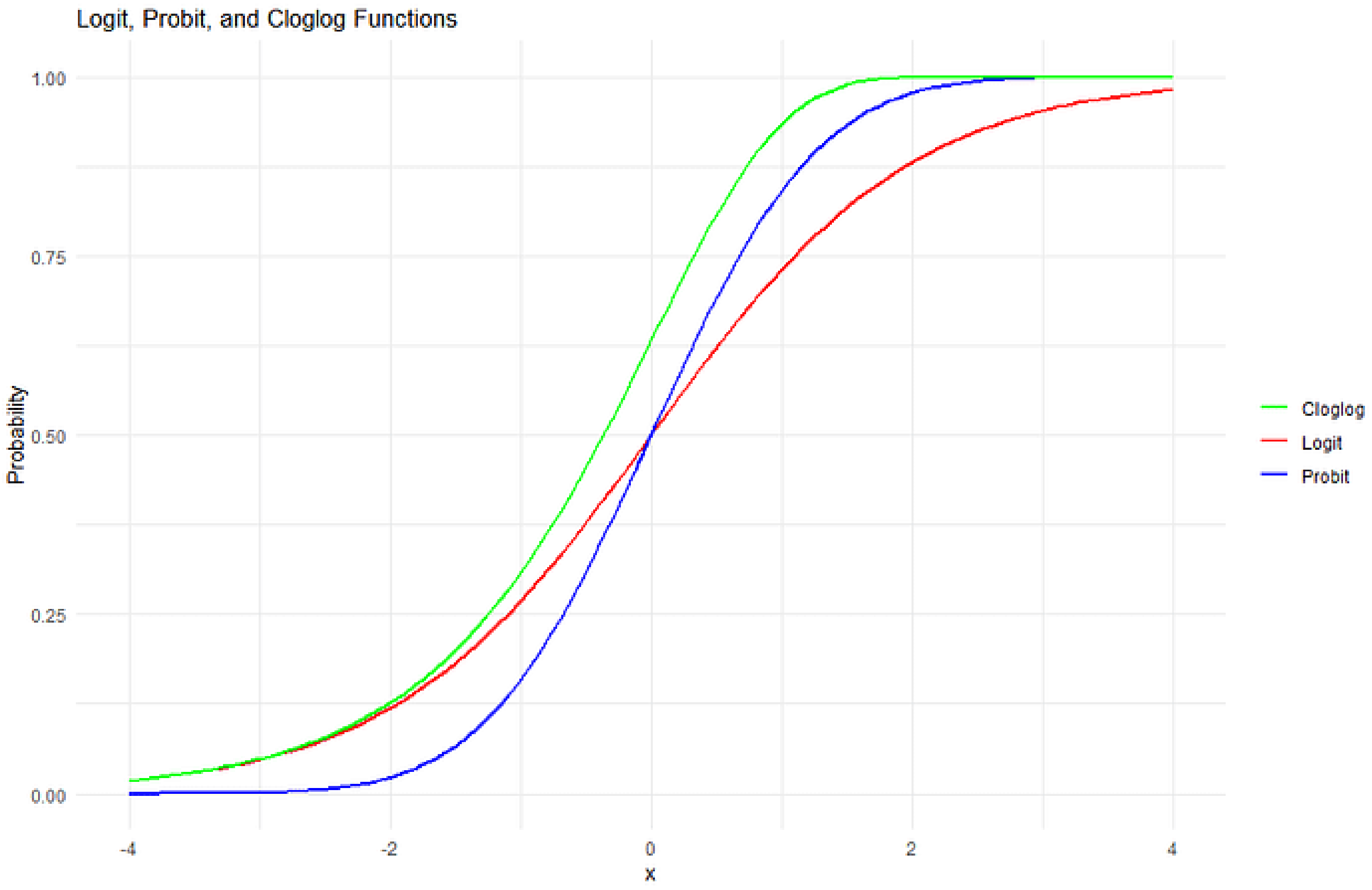
Different link functions (an example to constrain b/w 0 and 3)



$$y = 3 \times \frac{e^{\beta_0+\beta_1\times x}}{1+e^{\beta_0+\beta_1\times x}}$$

$$\beta_0 + \beta_1 \times x = \ln\left(\frac{\frac{y}{3}}{1-\frac{y}{3}}\right)$$

Different link functions (different shapes)



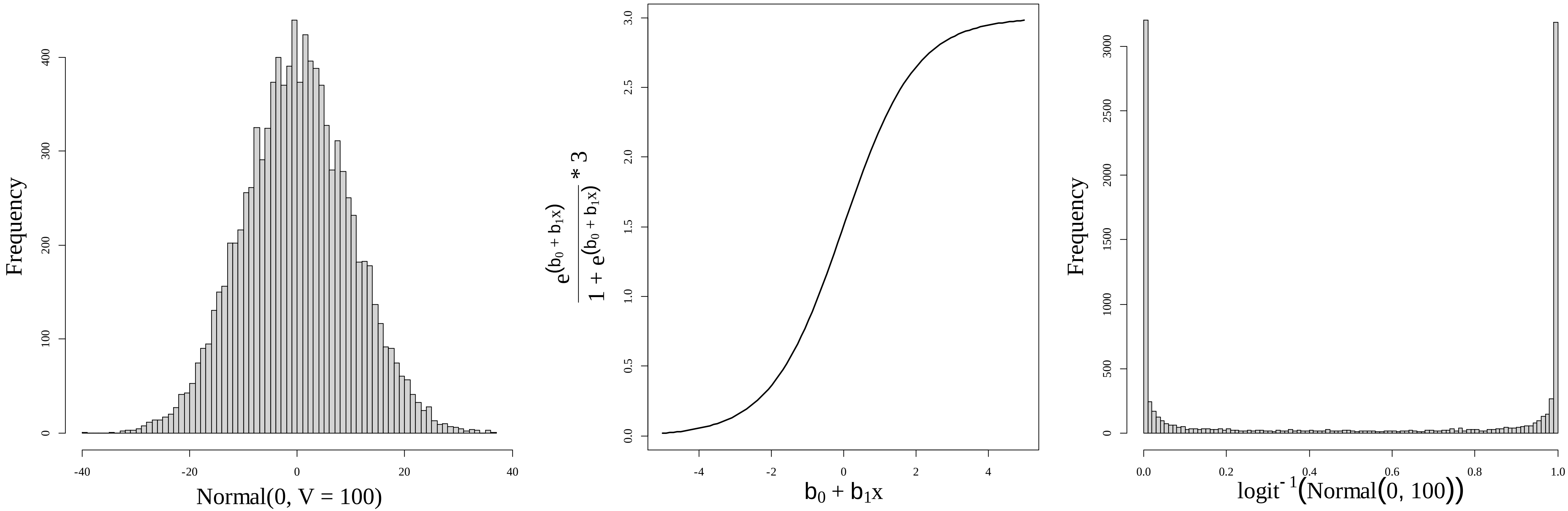
Things to think about:

1. Beta estimates are no longer ‘cut and dry’ rise/run. They are relative to your position on the link function and the shape of the link function.

2. There are many, many different types of link functions that do very similar things. Sometimes using a slightly different function can help with parameter estimation (related to point #1).

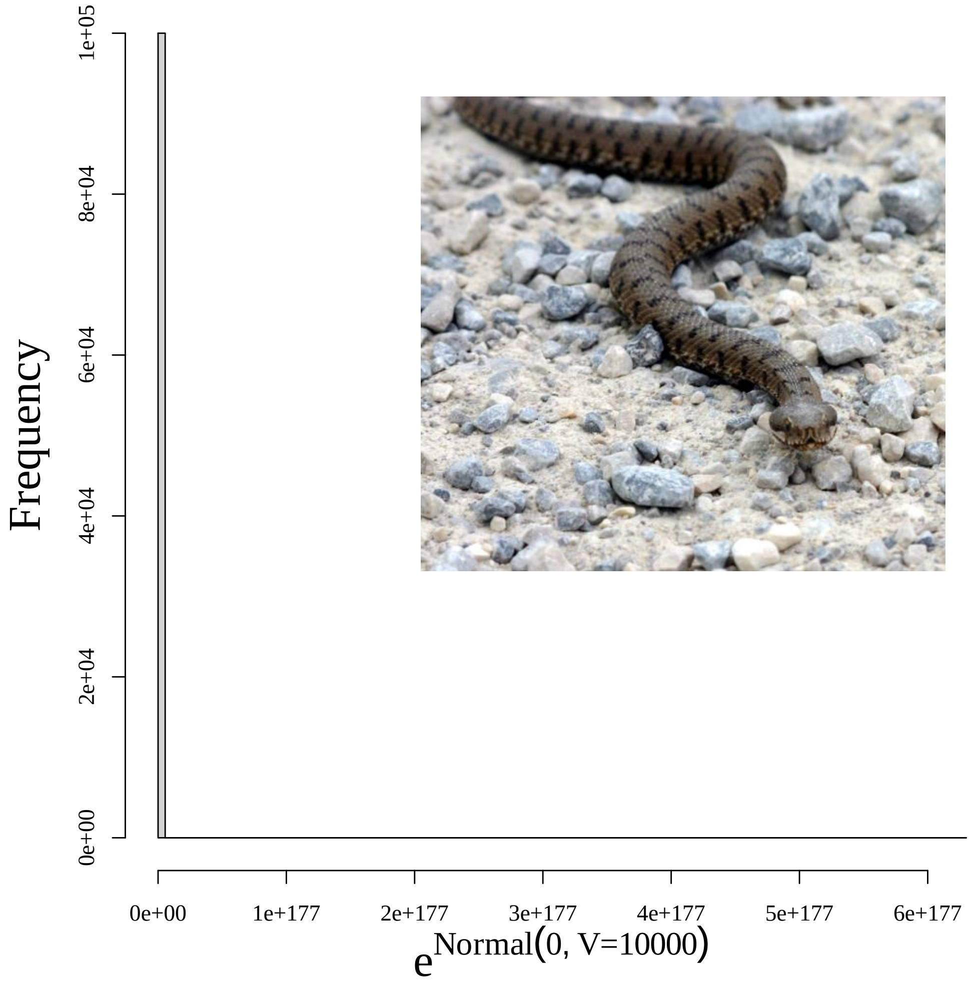
3. Priors suddenly become **quite a bit more complex** to think about... (see Northrup and Gerber 2018 for a logit-link example... and Lecture 2 for some log-link examples, e.g., snakes larger than a solar system!)

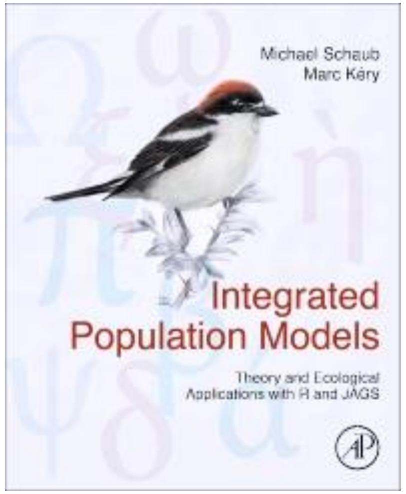
Priors on link functions



Northrup & Gerber **(2018)** *PLOS ONE*

Priors on link functions





Earth’s diameter is 12,756 km