

# Introduction to Bayesian Statistics

*Part 5*  
Generalized Linear Models

Benjamin Rosenbaum



iDiv 2025

# This lecture

Introduction to GLM

Discrete responses

- Logistic regression
- Binomial regression
- Poisson regression

Continuous responses

- Beta regression
- Distributional model

Nonlinear models

Summary

# Assumptions for linear models

1. Independent observations.

Systematic differences in  $y$  are because of  $x$  !

2. Trend of  $y$  follows (linear) prediction model

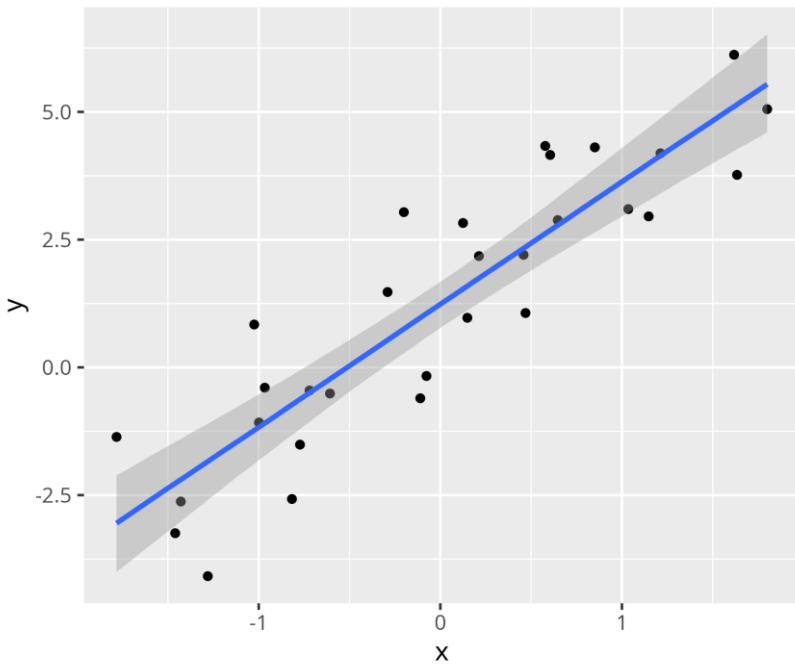
$$\mu(x) = a + b \cdot x$$

3. Residuals follow normal distribution

$$\varepsilon \sim \text{Normal}(0, \sigma)$$

4. Constant variance (standard deviation  $\sigma$ )

across whole range of  $x$



# Assumptions for linear models

1. Independent observations.

Systematic differences in  $y$  are because of  $x$  !

2. Trend of  $y$  follows (linear) prediction model

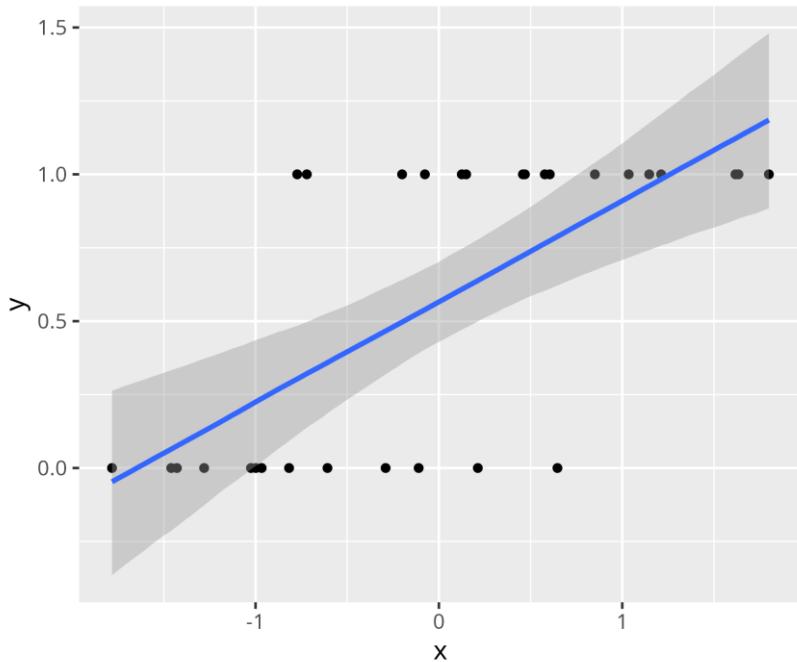
$$\mu(x) = a + b \cdot x$$

3. Residuals follow normal distribution

$$\varepsilon \sim \text{Normal}(0, \sigma)$$

4. Constant variance (standard deviation  $\sigma$ )

across whole range of  $x$



# What do we need?

**Deterministic part:**

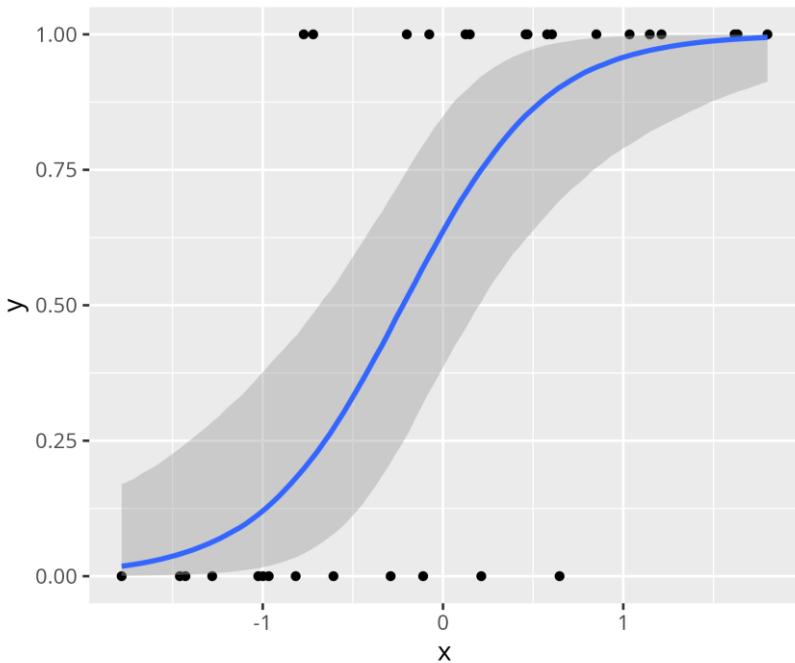
$$\mu(x) = a + b \cdot x$$

→ Regression curve  $\mu(x)$  that respects  
actual boundaries of the data

**Stochastic part:**

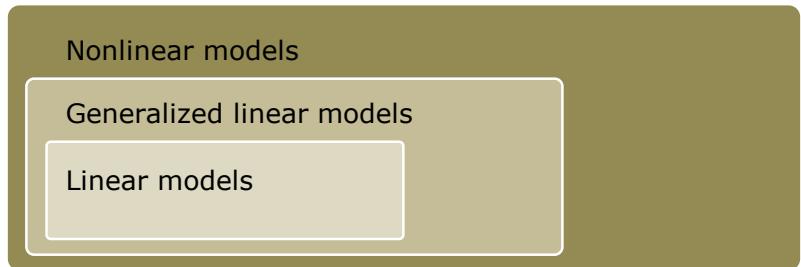
$$y \sim \text{Normal}(\mu, \sigma)$$

→ A distribution that models non-normal residuals  
correctly (discrete, skewed, non-constant var, ...)

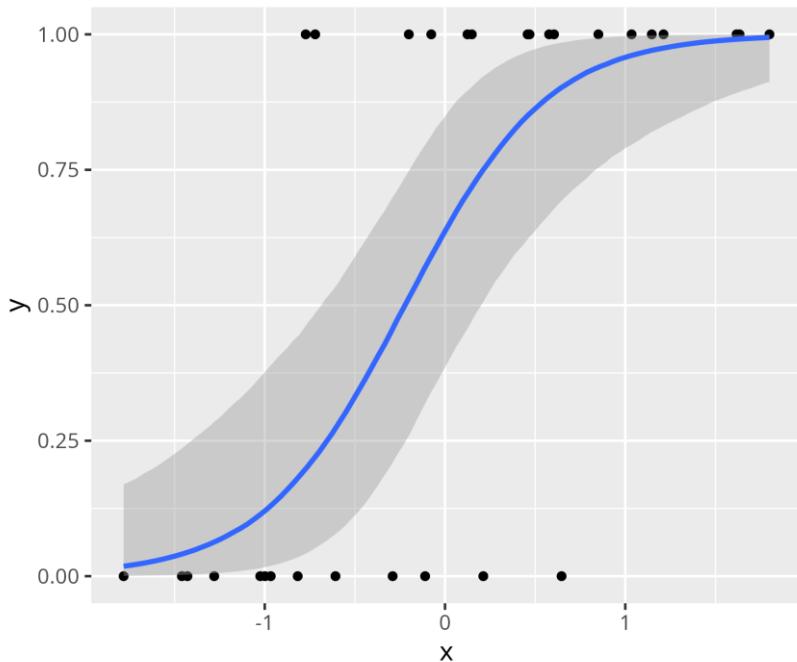


# The generalized linear model (GLM)

Theoretically, we could fit all kinds of nonlinear regression curves  $\mu(x)$  to the data



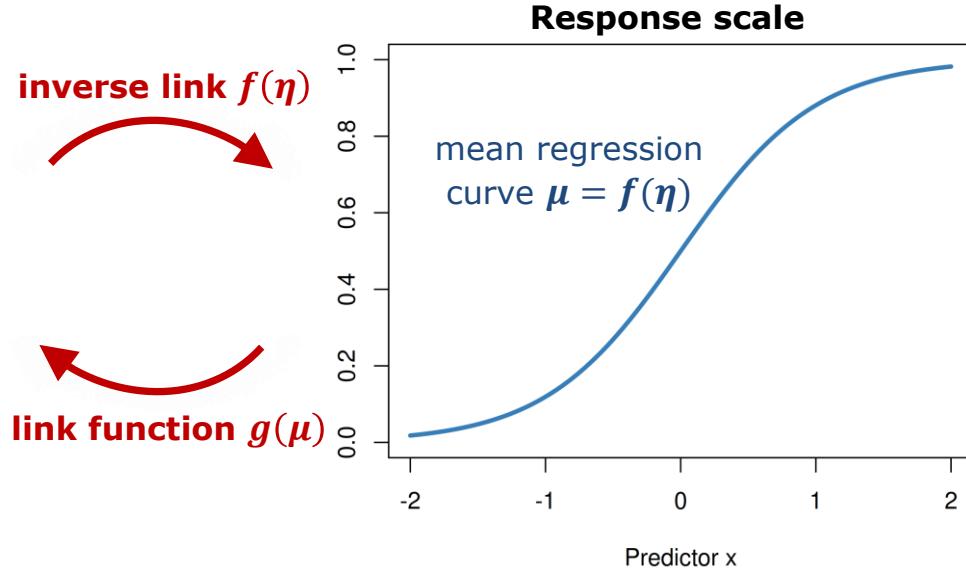
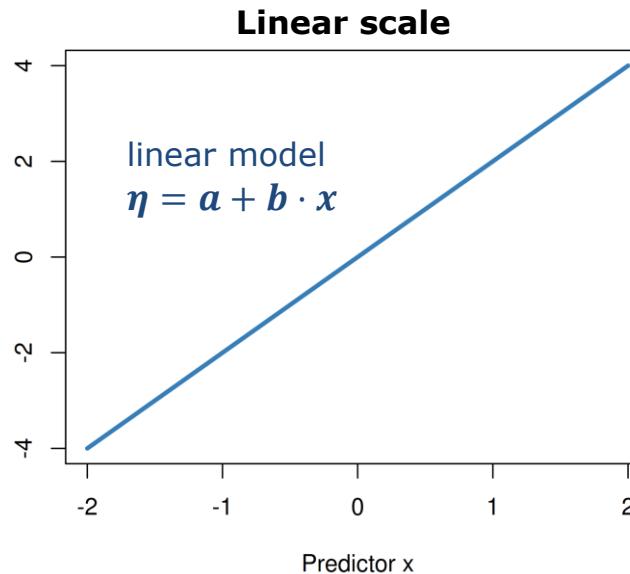
GLM is a special case of nonlinear models,  
that maps a linear model  $\eta(x) = a + bx$   
with a nonlinear function  $\mu = f(\eta(x))$   
to the response scale  $y$



Why? „Easy“ to use.

Model structures, interactions, effect sizes, etc in the linear part are meaningful

# Deterministic part: the link function $g$



Inverse link *compresses* linear model  $\eta(x)$  to response scale

Link function *expands* mean regression curve  $\mu(x)$  to linear scale

(Yes, this notation is very confusing)

# Stochastic part: a (non-normal) distribution

Linear model:  $\eta(x) = a + b \cdot x$

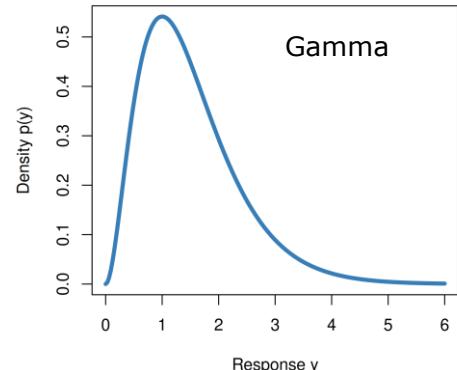
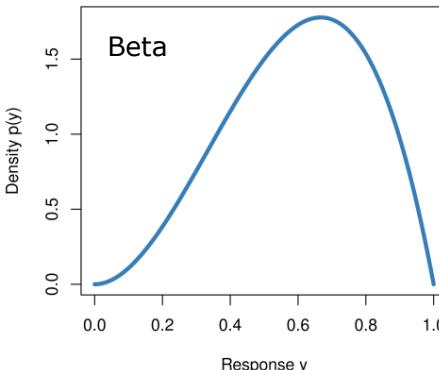
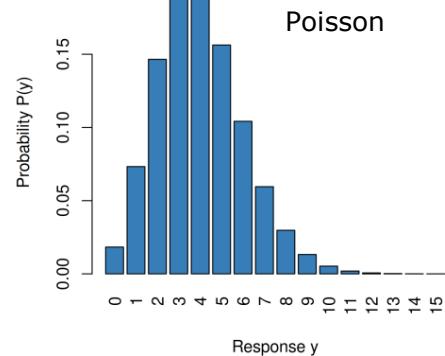
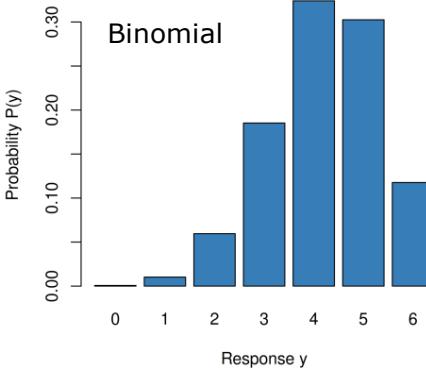
(Inverse) link:  $\mu = f(\eta)$

**Residual:**  $y \sim \text{Distribution}(\mu, \dots)$

mean prediction  
additional scale or dispersion parameters

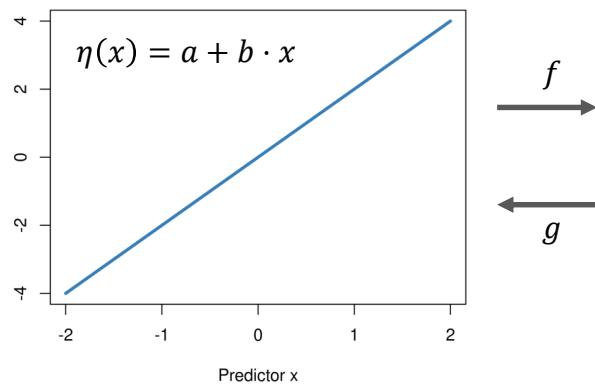
Distribution limited by type of data  $y$ :

- discrete or continuous ?
- constrained (lower/upper) or unconstrained ?
- skewed or symmetric ?
- mean-variance relationship ?

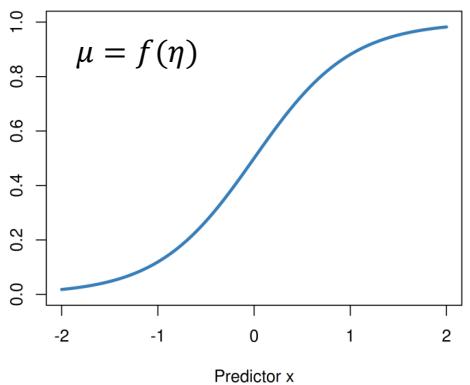


# 3 parts of the GLM

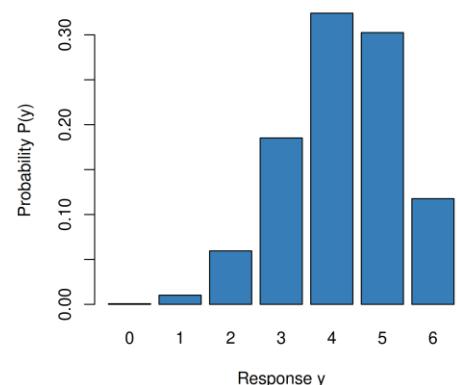
## 1: linear model



## 2: (inverse) link



## 3: residual distribution



- Which predictors ?
- How? Interactions, etc. ?
- All linear models allowed  
(also ANOVA, ANCOVA, ...)
- Assign **priors  $p(\theta)$**  for model parameters !

- Mean prediction  $\mu(x)$
- On response scale
- $f$  inverse link function
- $g$  link function

- How is  $y$  distributed around mean prediction  $\mu$  ?
- Defines **likelihood  $p(y|\theta)$**

# Computation

## Logical order of computation

$$\eta = a + b \cdot x$$

$$\mu = f(\eta)$$

$y \sim \text{Distribution}(\mu)$

Uses inverse-link  $f$

## Pseudo-code for Stan / Nimble / JAGS

```
for(i=1:n){  
  mu[i] = f(a+b*x[i])  
  y[i] ~ Distribution(mu[i])  
}
```

## Formula-based notation

$$g(\mu) = a + b \cdot x$$

$y \sim \text{Distribution}(\mu)$

Uses link function  $g$

## Base-R / rstanarm / brms:

```
glm( y~x, family=Distribution(link=g) )  
  
brm( y~x, family=Distribution(link=g),  
      prior=my.priors )
```

# *Logistic regression*

# Logistic regression

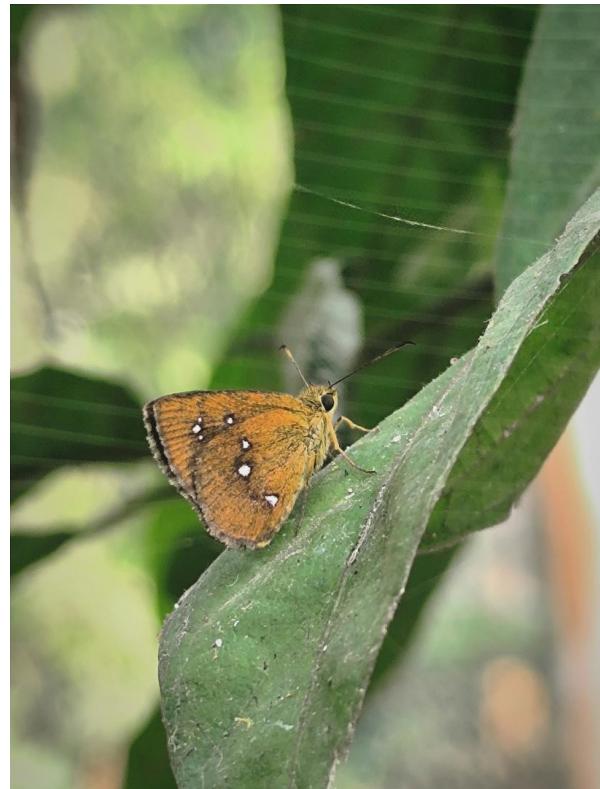
**Example:** Occurrence of a butterfly species  
versus temperature

Each observation is a habitat.

Dataset records presence / absence (1/0) of a rare species.

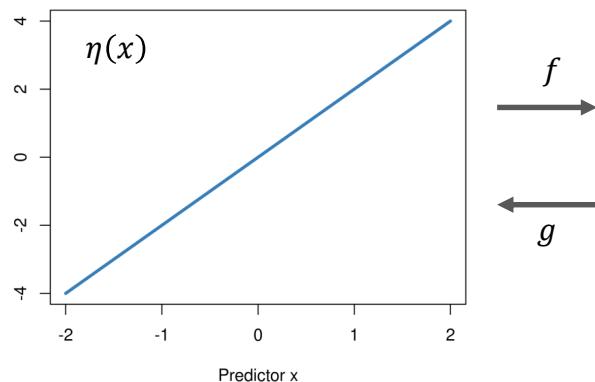
Mean annual temperature of each habitat as predictor.

**Q:** What is the relationship between temp. and occurrence?  
What is the expected probability of occurrence for a new  
habitat with a given temperature?

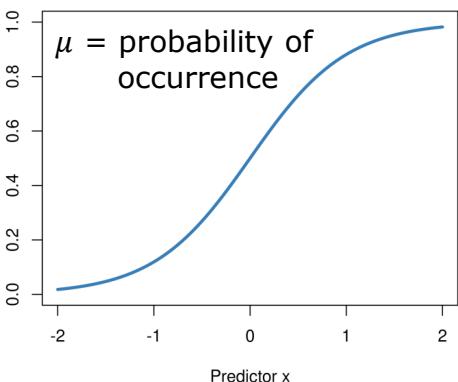


# Logistic regression: overview

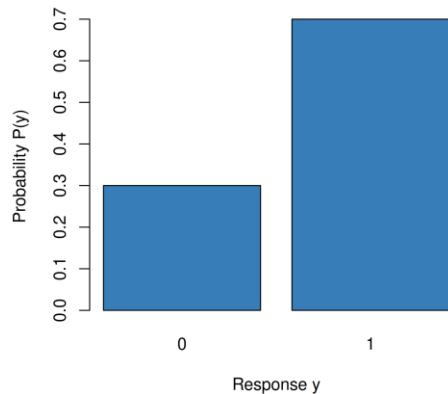
## 1: linear model



## 2: (inverse) link



## 3: residual distribution



1 predictor temperature:

$$\eta = b_0 + b_1 \cdot \text{temp}$$

## Logit link

$$g = \text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$$

Inverse-logit ("logistic")

$$f = \text{logistic}(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$$

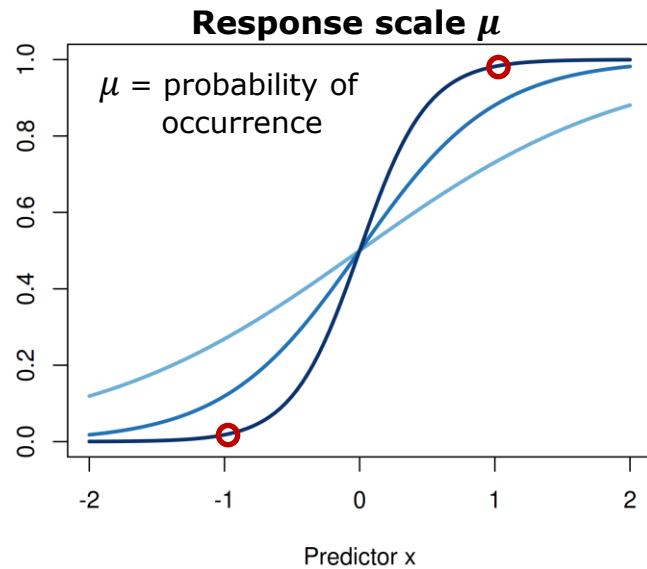
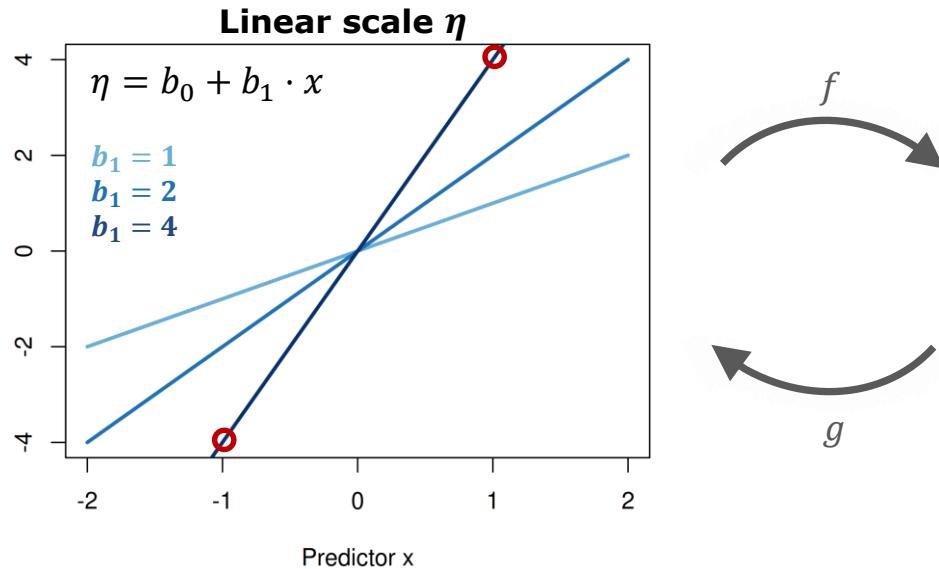
## Bernoulli distribution

$$y \sim \text{Bernoulli}(p = \mu)$$

$$= \begin{cases} p & (y = 1) \\ 1 - p & (y = 0) \end{cases}$$

Special case of Binomial distr.  
for  $N = 1$  trials

# Logistic regression: priors



On response scale:  $\mu(\eta = +4) = 0.982$  "almost always"  
 $\mu(\eta = -4) = 0.018$  "almost never"

→ Scale predictors  $x$  & choose meaningful prior for slopes, e.g.  $b_1 \sim \text{Normal}(0,1)$

# Logistic regression: model fit

**Deterministic part**  
(linear model & link)

$$\text{logit}(p) = b_0 + b_1 \cdot \text{temp}$$

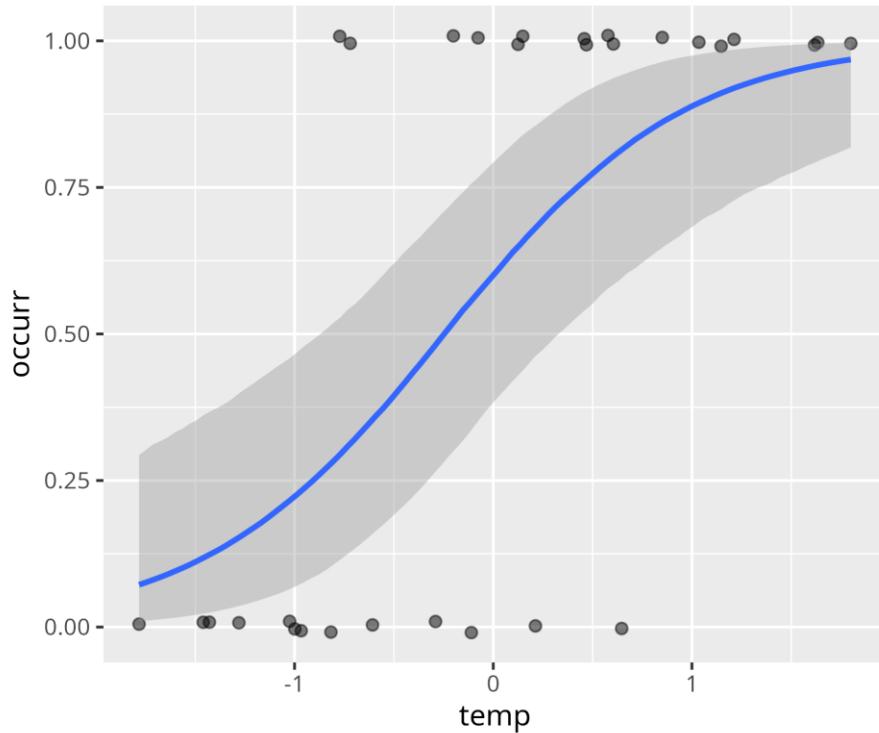
**Stochastic part**

$$\text{occur} \sim \text{Bernoulli}(p)$$

**Priors**

$$\begin{aligned} b_0 &\sim \text{brms-default} \\ b_1 &\sim \text{Normal}(0,1) \end{aligned}$$

```
> brm(occur~temp, family=bernoulli(link=logit),  
      prior =... )
```



# Logistic regression: model fit

```
> brm(occur~temp, family=bernoulli(link=logit),  
      prior =... )
```

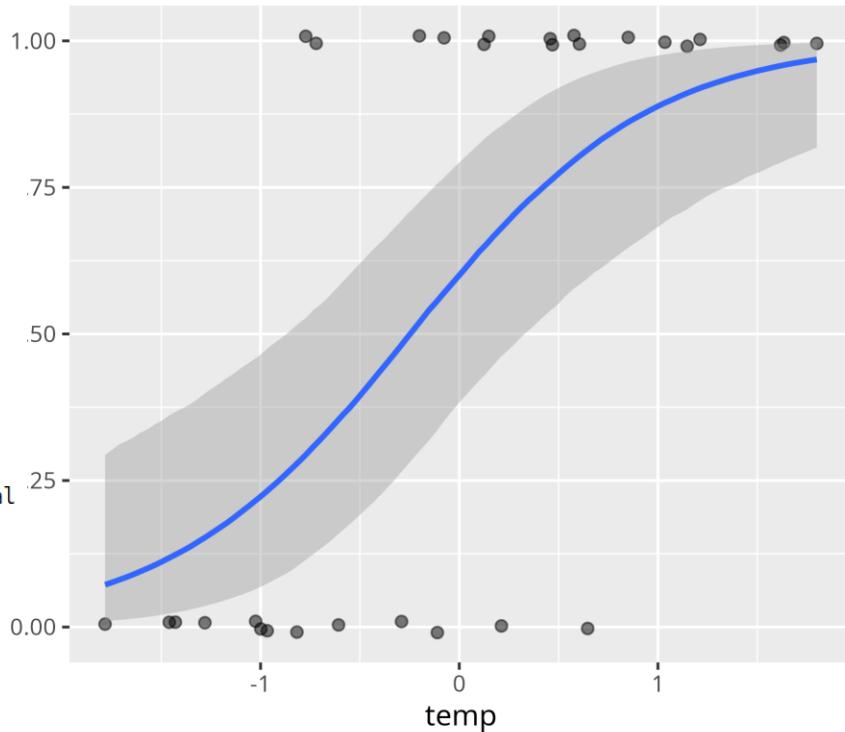
```
Family: bernoulli  
Links: mu = logit  
Formula: occur ~ temp  
Data: df (Number of observations: 30)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
      total post-warmup draws = 4000
```

```
Regression Coefficients:  
Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
Intercept 0.41 0.46 -0.48 1.34 1.00 3132 2525  
temp 1.70 0.51 0.77 2.78 1.00 2638 2078
```

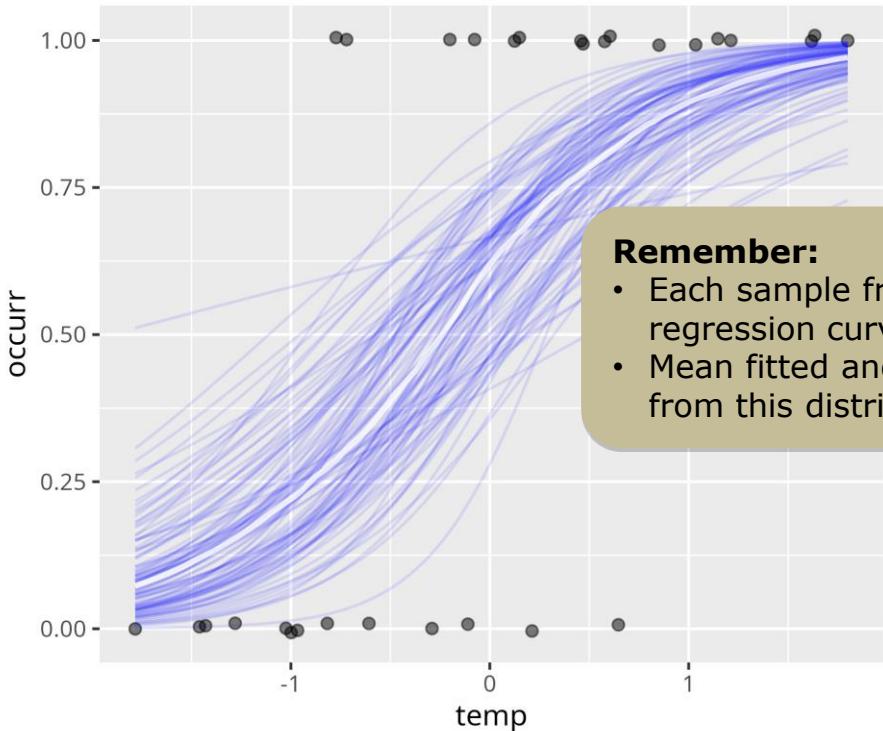
Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS and Tail\_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

**Q:** Is there a positive relationship between temperature and occurrence?

→ Yes, positive:  $b_1 = 1.70 [0.77, 2.78]$

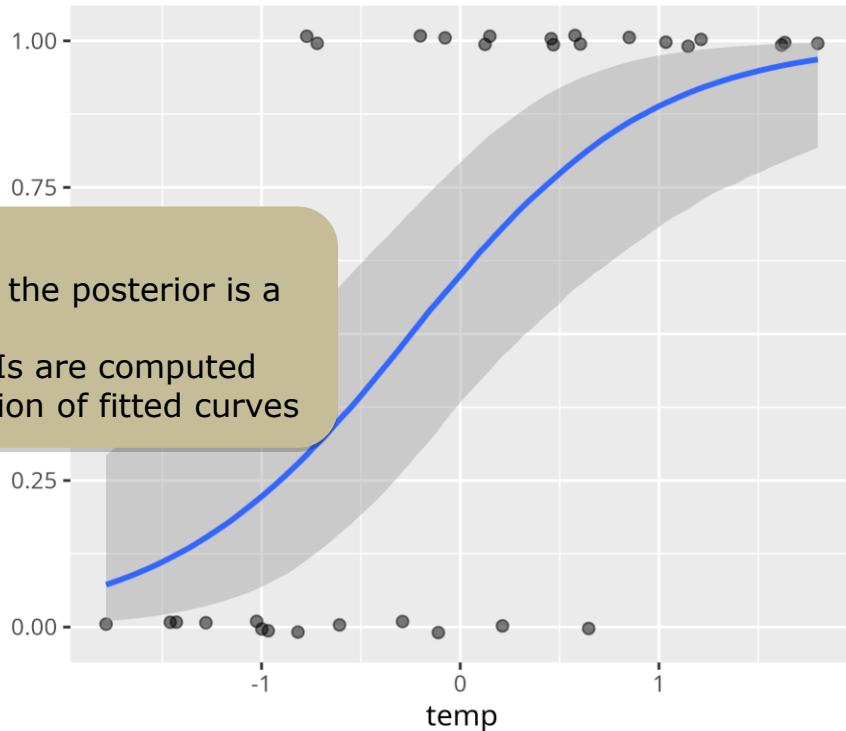


# Logistic regression: model fit

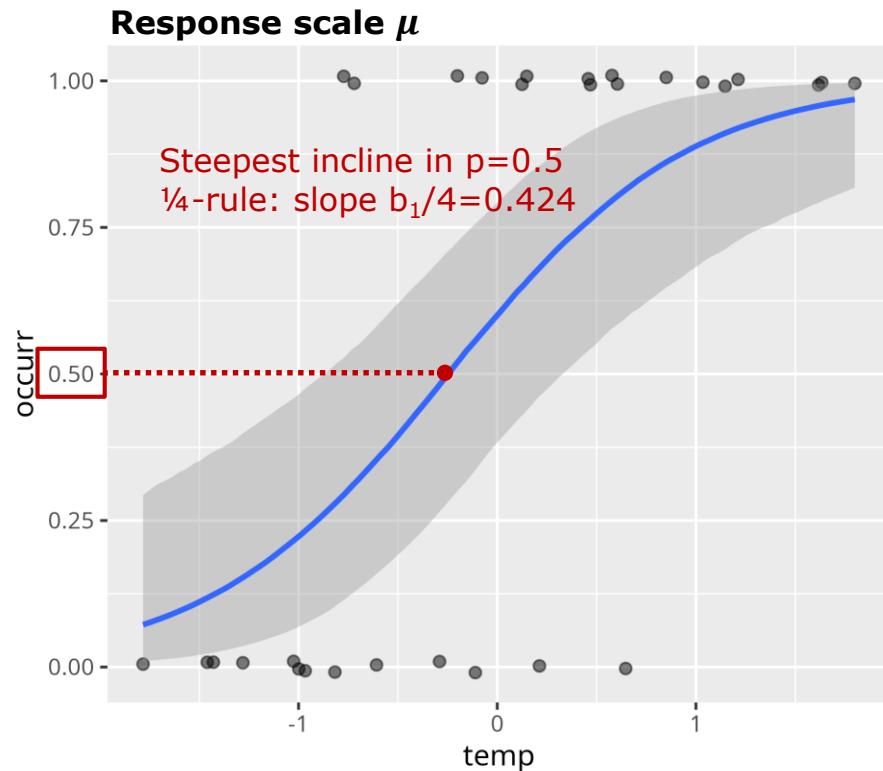
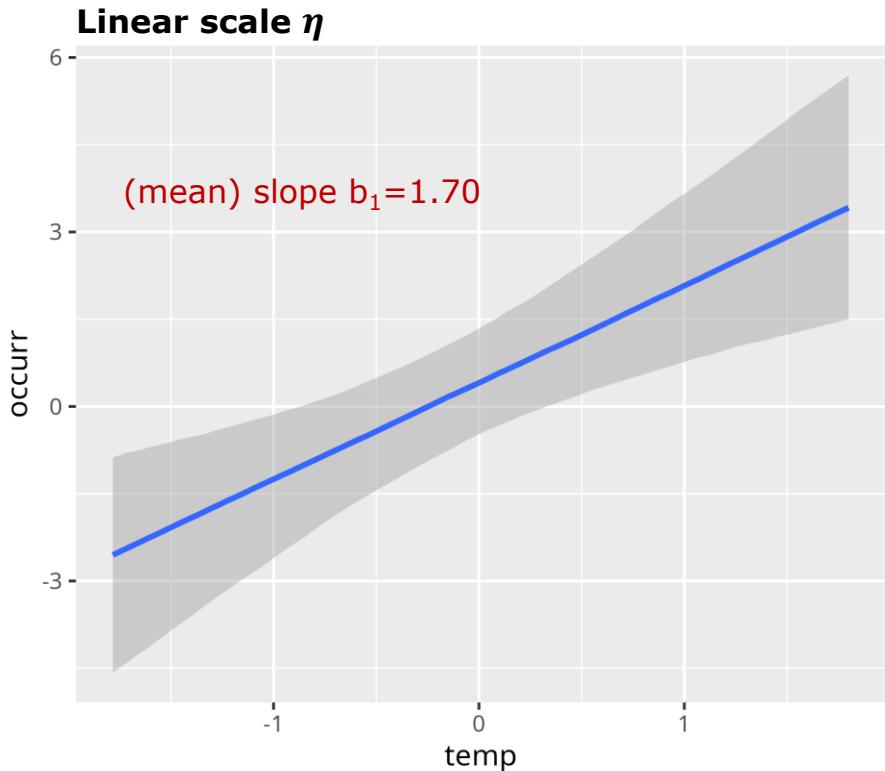


**Remember:**

- Each sample from the posterior is a regression curve
- Mean fitted and CIs are computed from this distribution of fitted curves

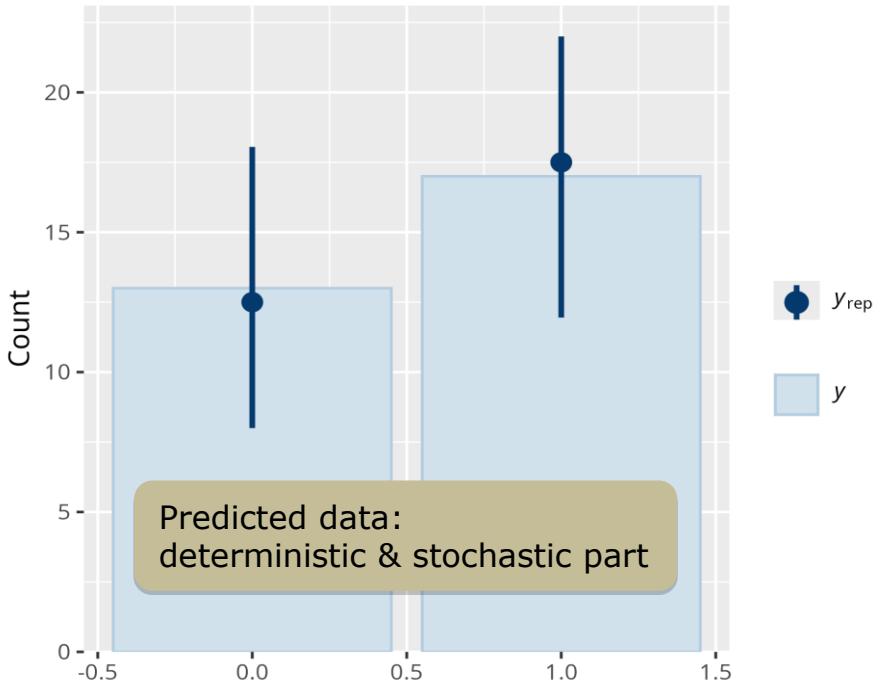


# Logistic regression: model fit

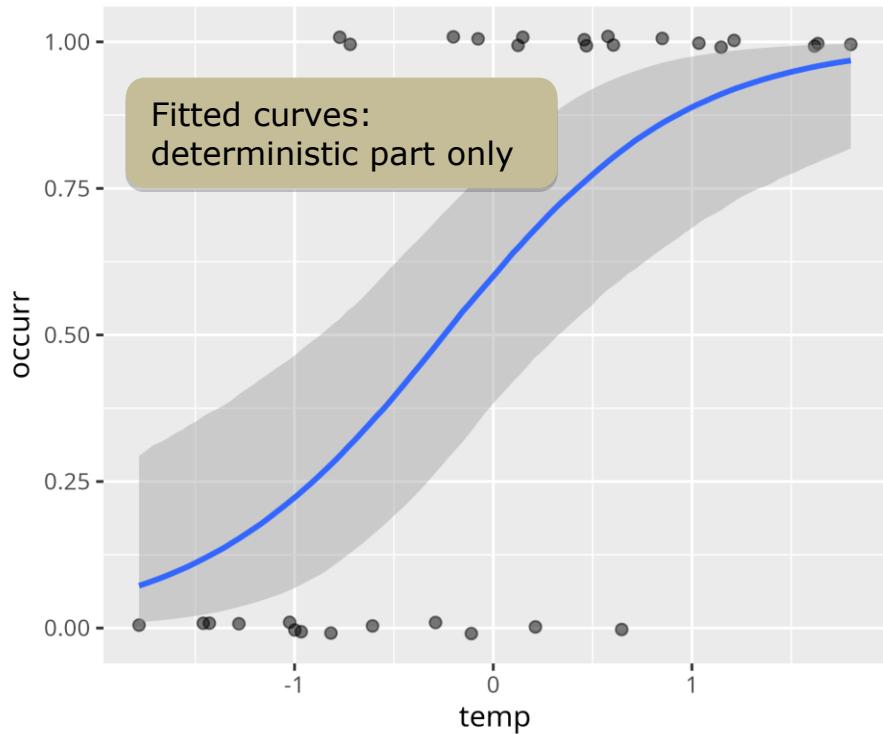


# Logistic regression: evaluation

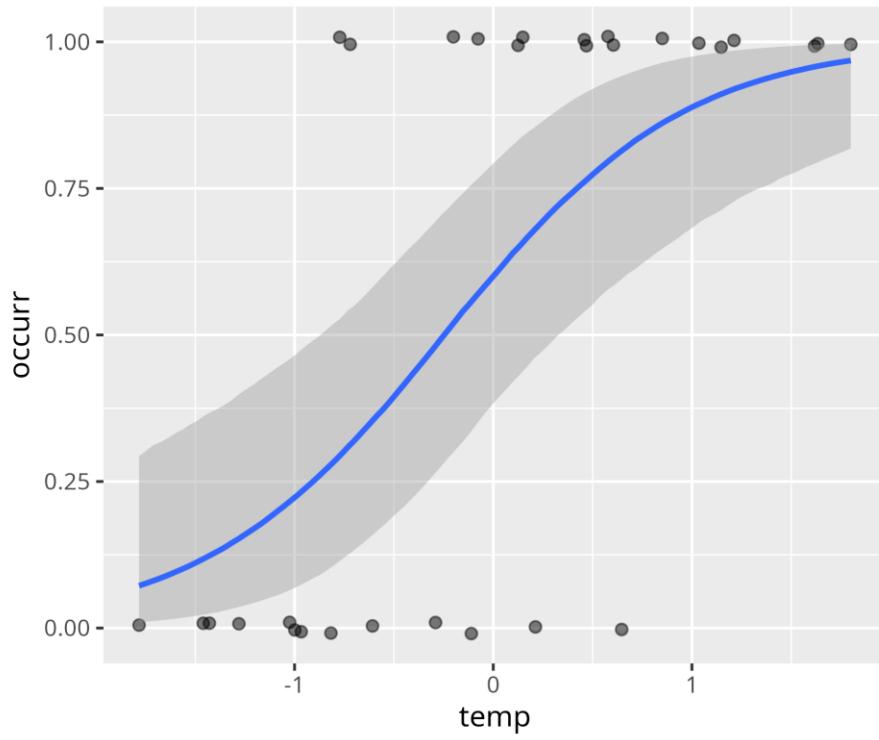
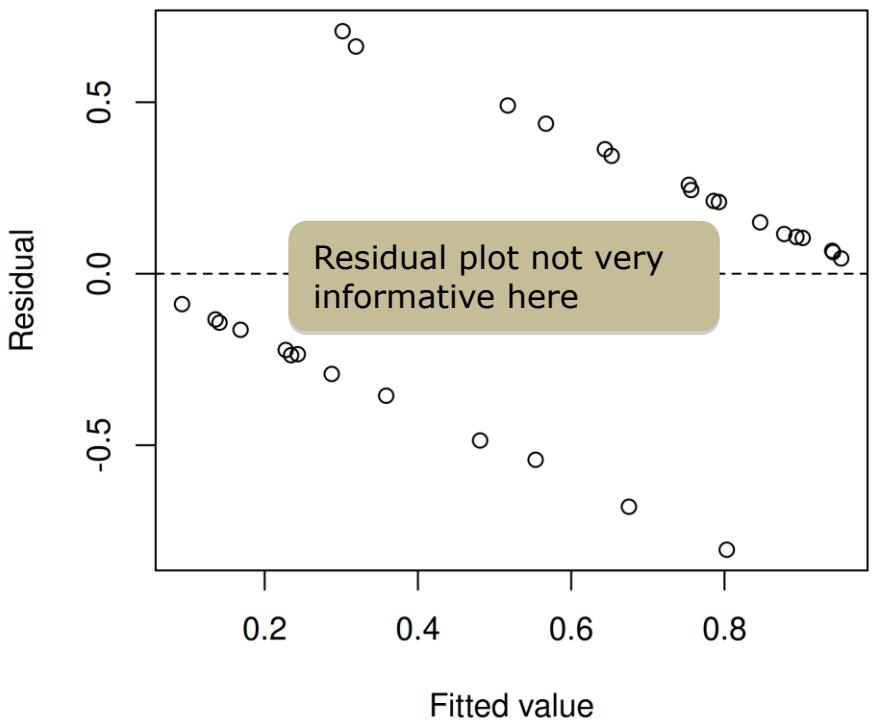
```
> pp_check(fit, type=„bars“)
```



```
> plot(conditional_effects(fit,...))
```

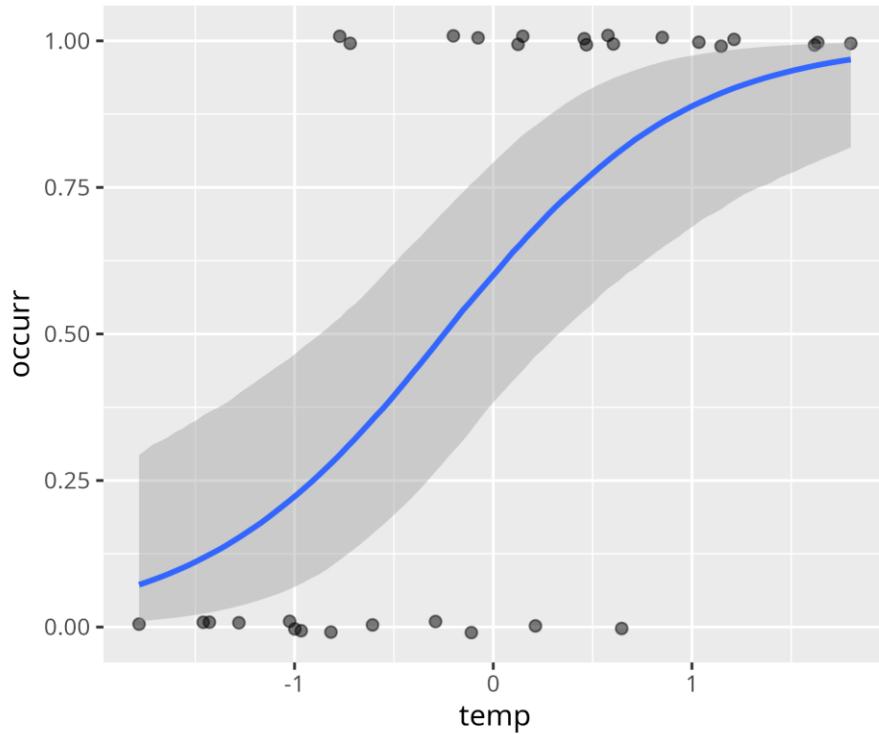
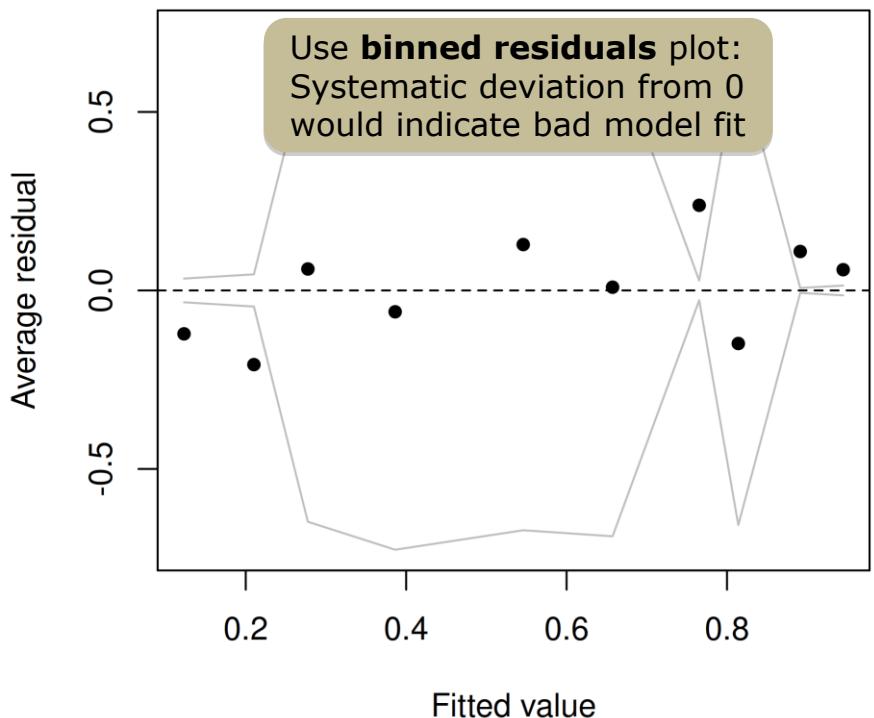


# Logistic regression: evaluation



# Logistic regression: evaluation

```
> arm:::binnedplot(fitted, residuals)
```



# Logistic regression (Niche model)

**Example:** Occurrence of a butterfly species  
versus temperature

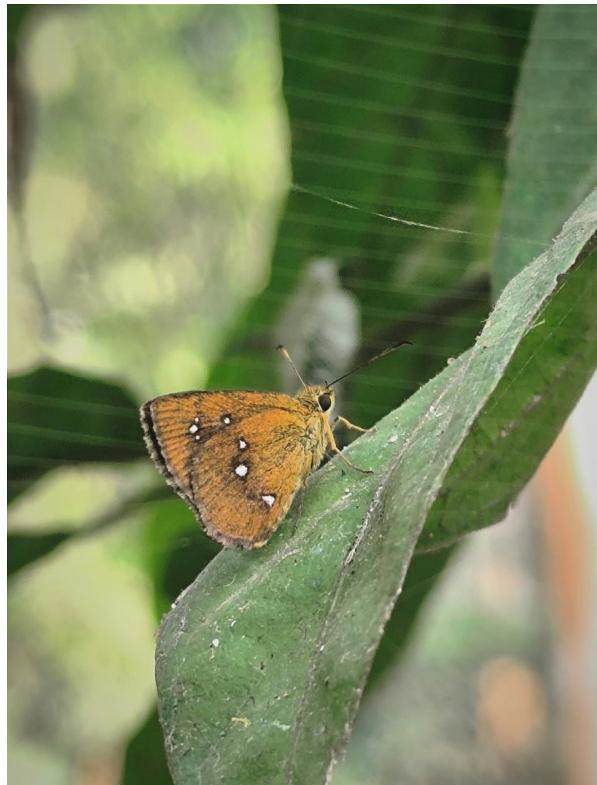
Each observation is a habitat.

Dataset records presence / absence (1/0) of a rare species.

Mean annual temperature of each habitat as predictor.

**New data collected in warmer habitats.**

**Q:** What is the optimal temperature for this species ?



# Logistic regression: model fit

## Deterministic part

(linear model & link)

$$\text{logit}(p) = b_0 + b_1 \cdot \text{temp} \\ + b_2 \cdot \text{temp}^2$$

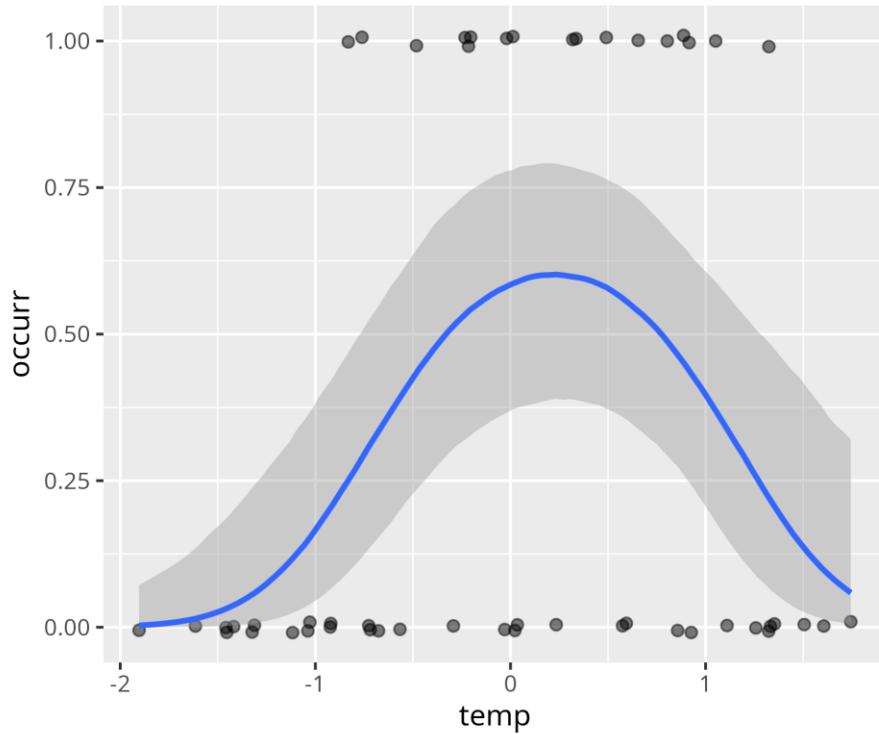
## Stochastic part

$$\text{occur} \sim \text{Bernoulli}(p)$$

## Priors

$$b_0 \sim \text{brms-default} \\ b_1 \sim \text{Normal}(0,1)$$

```
> brm(occur ~ temp + I(temp^2),  
      family = bernoulli(link=logit),  
      prior = ... )
```



# Logistic regression: model fit

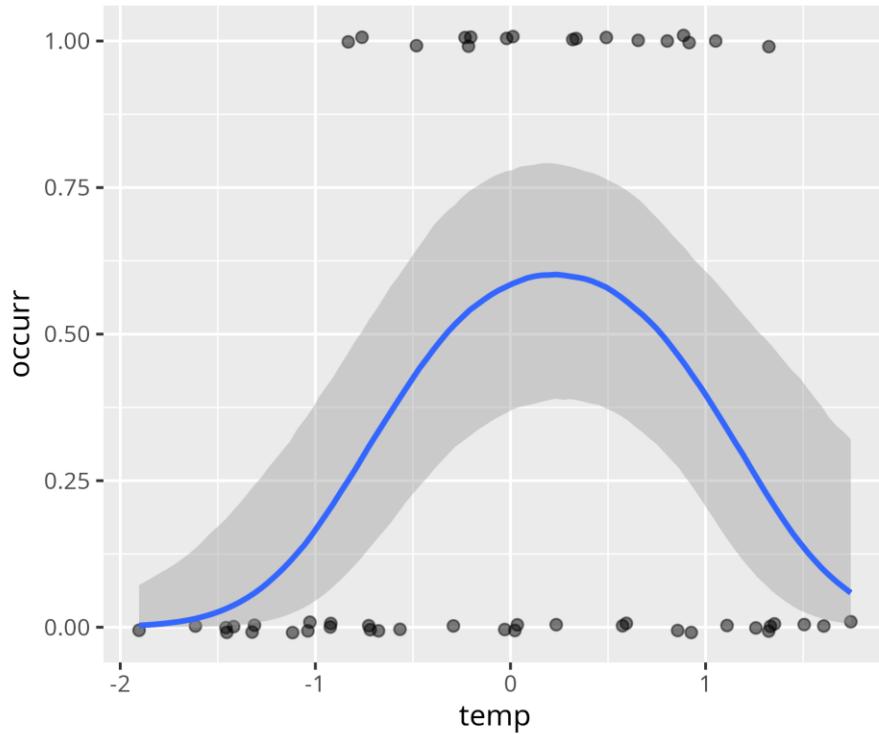
```
> brm(occurr ~ temp + I(temp^2),  
      family = bernoulli(link=logit),  
      prior  = ... )
```

```
Family: bernoulli  
Links: mu = logit  
Formula: occurr ~ temp + I(temp^2)  
Data: df (Number of observations: 50)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
      total post-warmup draws = 4000
```

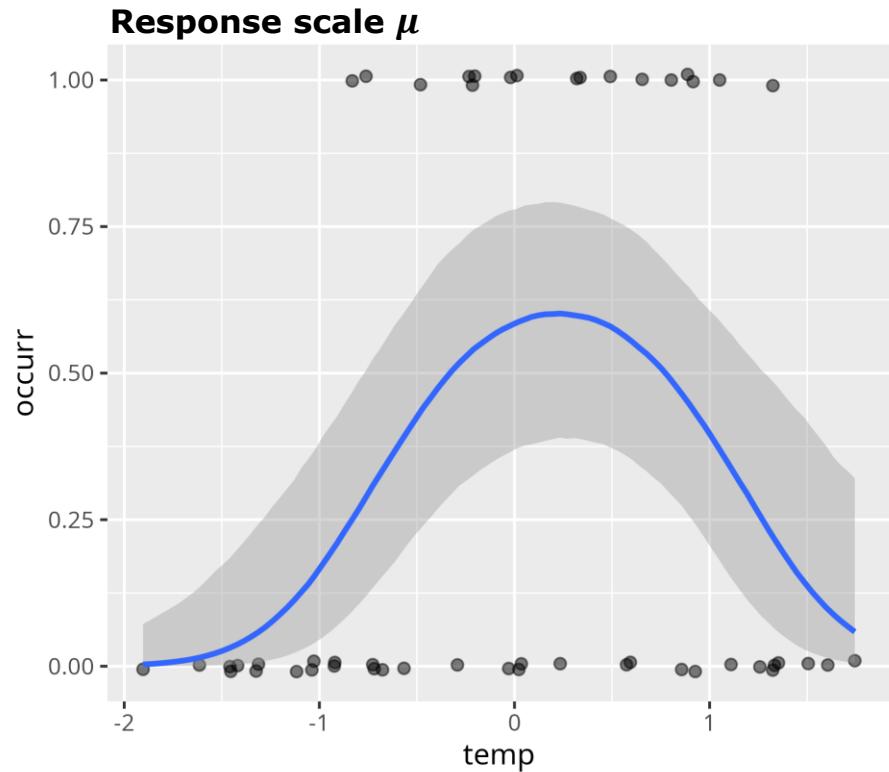
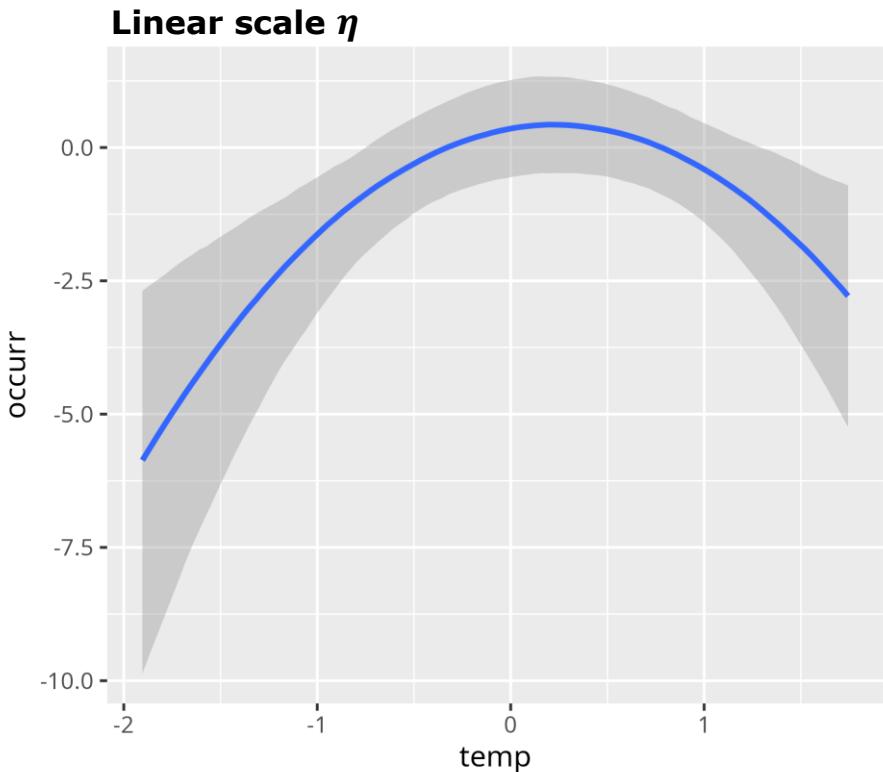
## Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.36	0.46	-0.55	1.26	1.00	5301	3161
temp	0.63	0.42	-0.15	1.49	1.00	2167	2154
ItempE2	-1.41	0.46	-2.36	-0.53	1.00	2031	2262

Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS and Tail\_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).



# Logistic regression: model fit

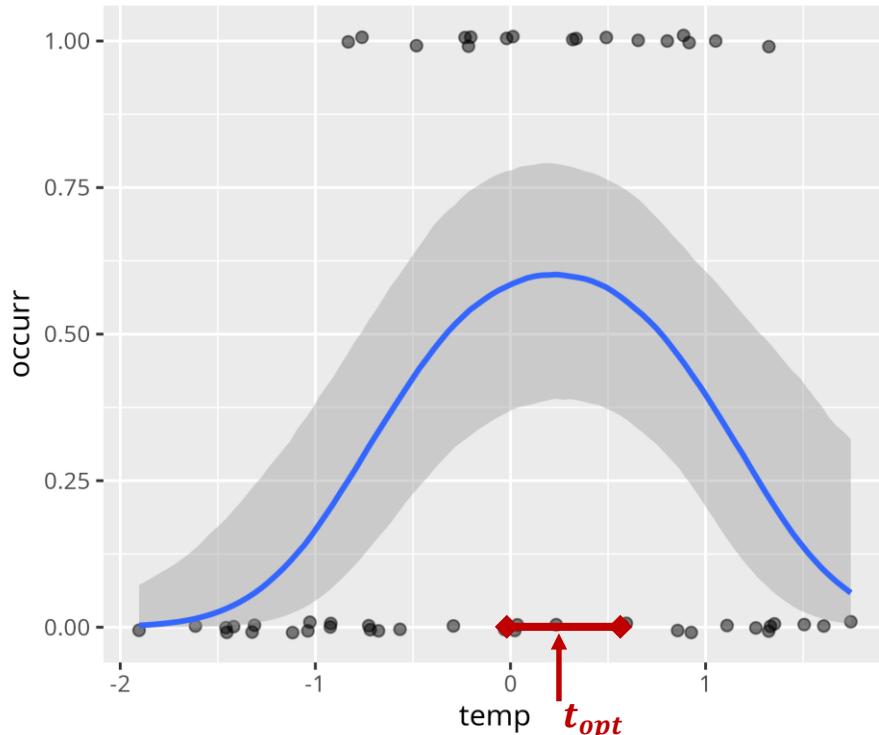
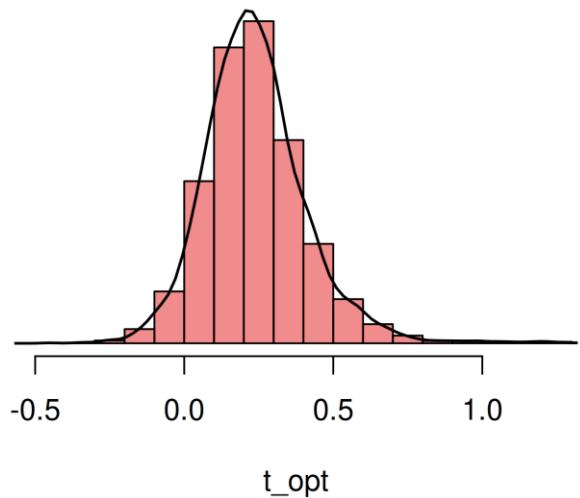


# Logistic regression: model fit

**Q:** What is the optimal temperature?  
(peak of the curve)

some calculus  $\rightarrow t_{\text{opt}} = -\frac{b_1}{2b_2}$

compute  $t_{\text{opt}}$  for each posterior sample  
 $\rightarrow$  mean 0.22, 90% CI [0.01, 0.54]





# Binomial regression

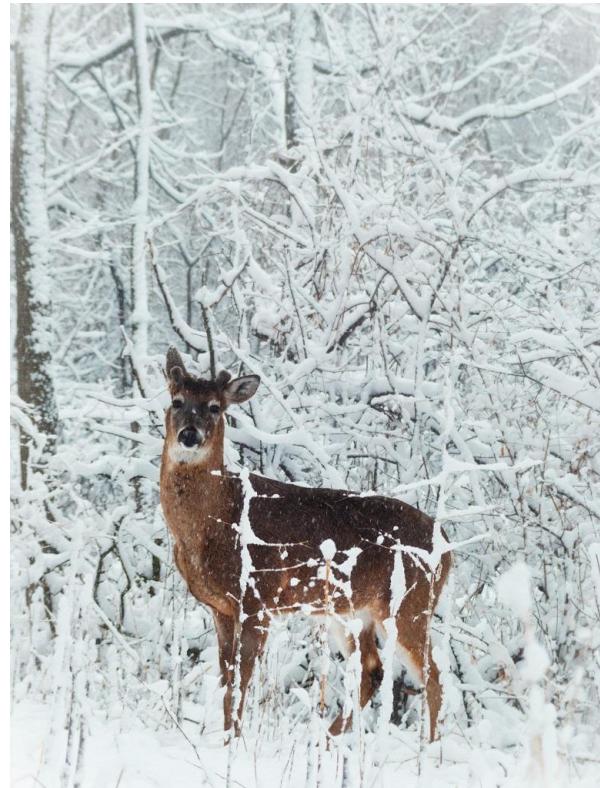
**Example:** Deer survival probability  
versus vegetation and predation

Monitored deer populations several habitats.

Recorded number of deer before and after winter.

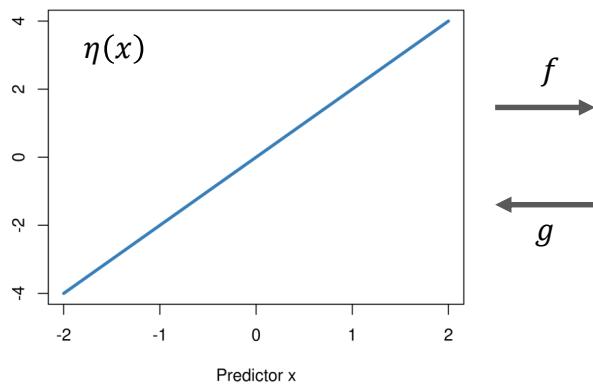
Vegetation index (NDVI) and predator presence (yes/no).

**Q:** How much does vegetation affect survival probability?  
Must control for predation. Is there an interaction?

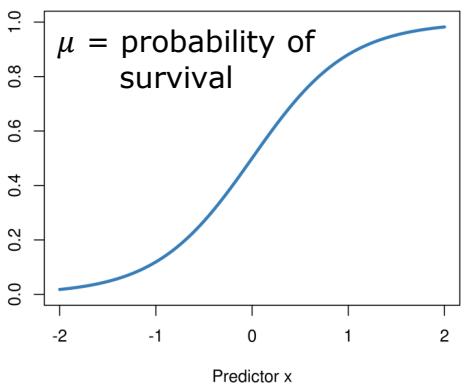


# Binomial regression: overview

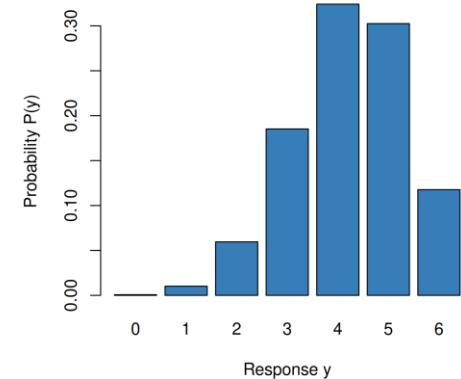
1: linear model



2: (inverse) link



3: residual distribution



## Logit link

$$g = \text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$$

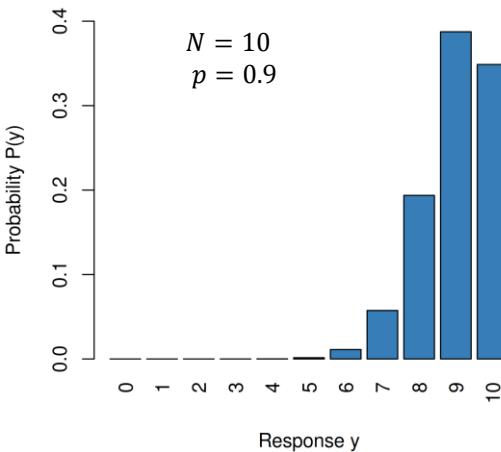
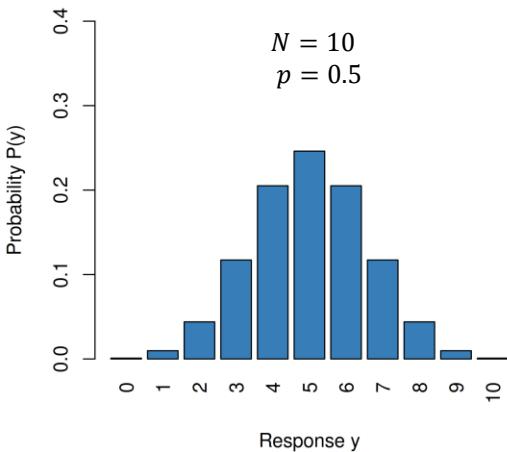
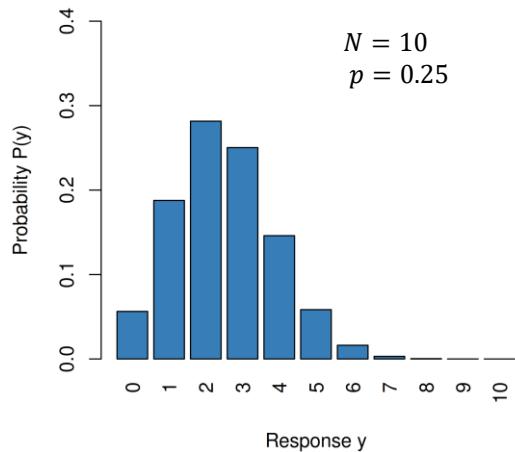
Inverse-logit ("logistic")

$$f = \text{logistic}(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$$

## Binomial distribution

$$y \sim \text{Binomial}(N, p = \mu)$$

# Binomial distribution



Discrete distribution, bounded between 0 and  $N$

Mean:  $N \cdot p$

Sdev:  $\sqrt{Np(1 - p)}$  (increases with  $N$ )

Every datapoint  $y_i$  can have its own  $N_i$

# Binomial regression: ANCOVA model

**Deterministic part**

(linear model &amp; link)

$$\text{logit}(p) = a_0 + a_1 \cdot x_{pred} + (b_0 + b_1 \cdot x_{pred}) \cdot \text{vegetation}$$

Dummy-coding

$$x_{pred} = \begin{cases} 0, & \text{predator} = no \\ 1, & \text{predator} = yes \end{cases}$$

**Stochastic part**

$$\text{survived} \sim \text{Binomial}(\text{total}, p)$$

**Priors**

$$a_0 \sim \text{brms-default}$$

$$a_1, b_0, b_1 \sim \text{Normal}(0, 1)$$

```
> brm(survived | trials(total) ~ vegetation * predator,
       family = binomial(link=logit),
       prior = ... )
```

	survived	total	vegetation	predator
1	7	40	0.2663940	1
2	21	25	0.7321366	0
3	5	15	0.6159937	1
4	3	37	0.2512467	1
5	26	28	0.9494554	1
6	18	20	0.9491275	0
7	4	14	0.2162431	0
8	26	27	0.8501039	0
9	26	35	0.5212167	0
10	6	11	0.5949854	0
11	14	16	0.5974067	0
12	4	13	0.3150053	1
13	33	39	0.7844620	1
14	2	22	0.2627381	1
15	4	20	0.4647540	1

# Binomial regression: model fit additive

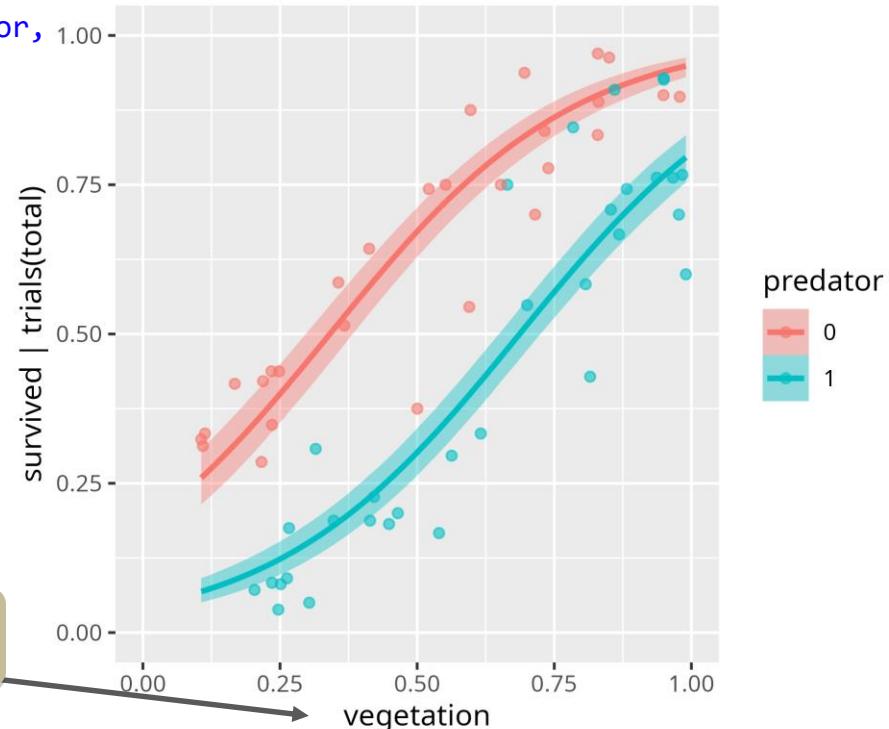
```
> brm(survived|trials(total) ~ vegetation+predator,  
      family = binomial(link=logit),  
      prior = ... )
```

```
Family: binomial  
Links: mu = logit  
Formula: survived | trials(total) ~ scale(vegetation) + predator  
Data: df (Number of observations: 60)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
      total post-warmup draws = 4000
```

```
Regression Coefficients:  
Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
Intercept     1.03     0.10    0.84    1.23 1.00   3104   3066  
scalevegetation 1.27     0.07    1.13    1.41 1.00   3207   2747  
predator1     -1.56     0.14   -1.83   -1.29 1.00   3046   2680
```

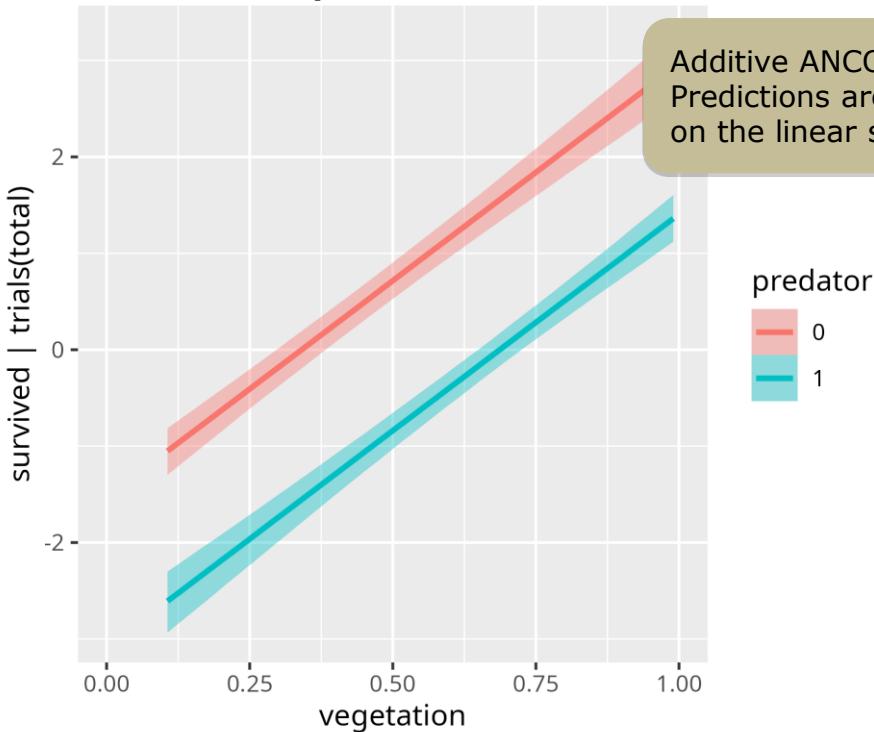
Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS and Tail\_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

Can also use `scale(...)` in the model formula.  
Predictions can be generated on original scale



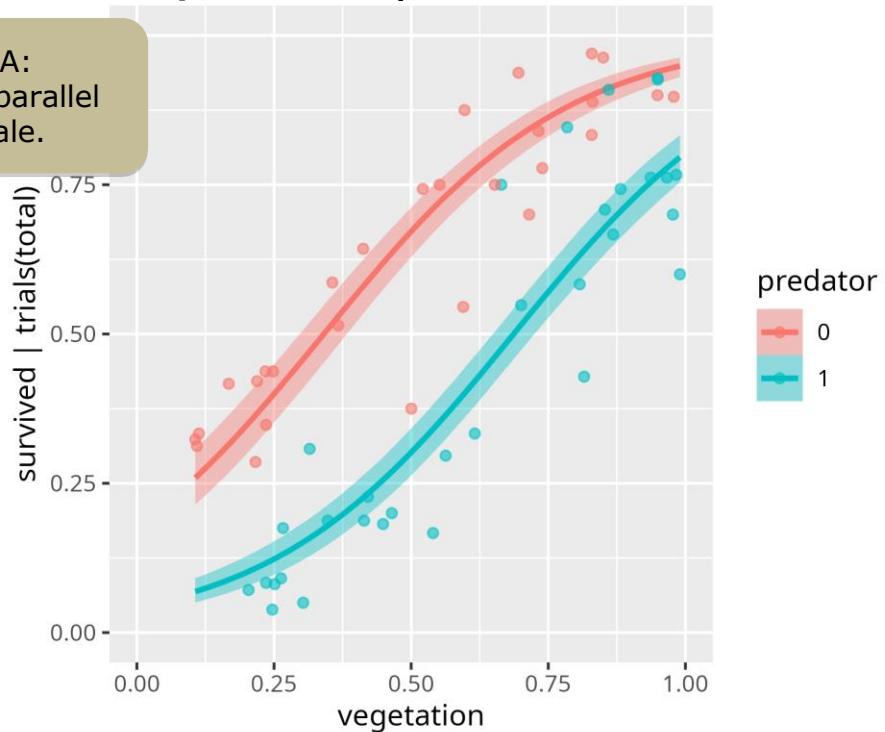
# Binomial regression: model fit additive

Linear scale  $\eta$



Additive ANCOVA:  
Predictions are parallel  
on the linear scale.

Response scale  $\mu$



# Binomial regression: model fit interaction

```
> brm(survived|trials(total) ~ vegetation*predator,
      family = binomial(link=logit),
      prior = ... )
```

Family: binomial  
Links: mu = logit  
Formula: survived | trials(total) ~ scale(vegetation) \* predator  
Data: df (Number of observations: 60)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
total post-warmup draws = 4000

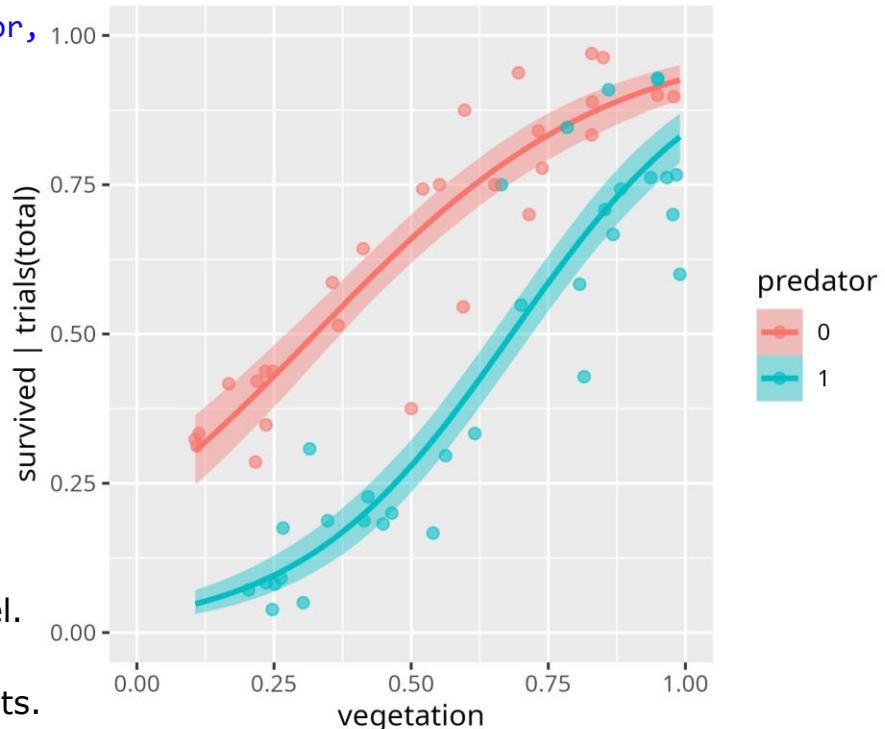
Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS
Intercept	0.93	0.10	0.73	1.14	1.00	2764
scalevegetation	1.07	0.10	0.88	1.27	1.00	2284
predator1	-1.51	0.14	-1.81	-1.23	1.00	3140
scalevegetation:predator1	0.39	0.14	0.12	0.67	1.00	2125

Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS and Tail\_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

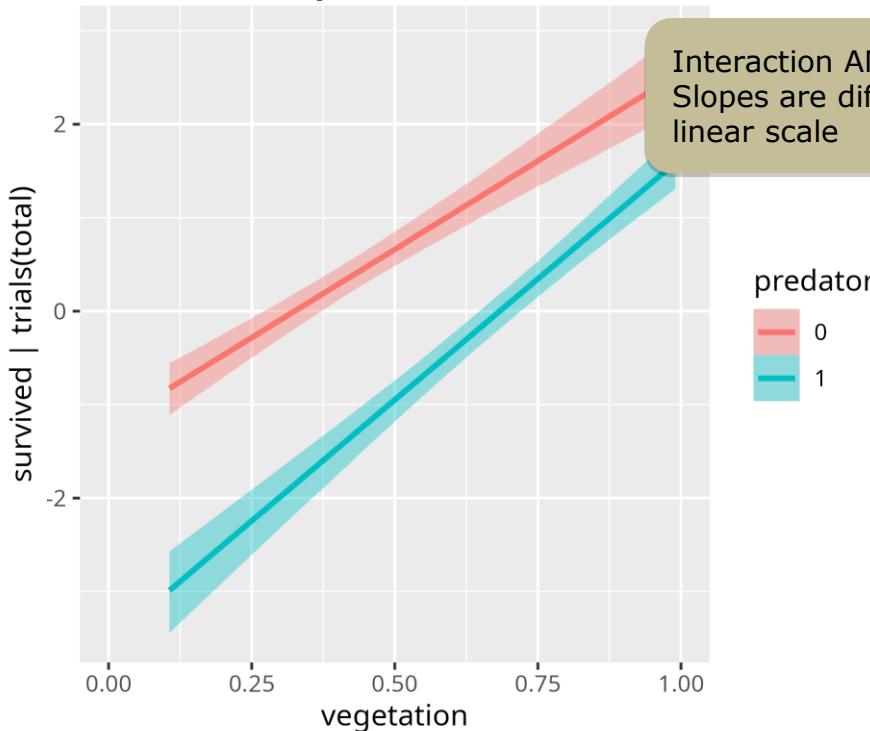
Alternative: model comparison against additive model.

→ Vegetation has a stronger effect in predator habitats.  
Important as shelter to hide from predators.

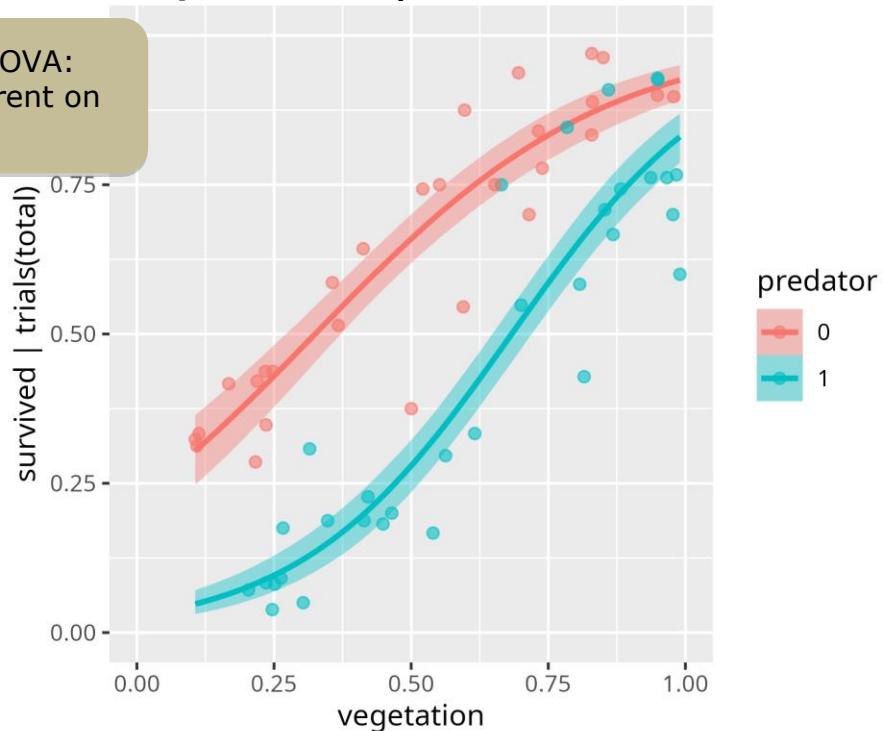


# Binomial regression: model fit interaction

Linear scale  $\eta$

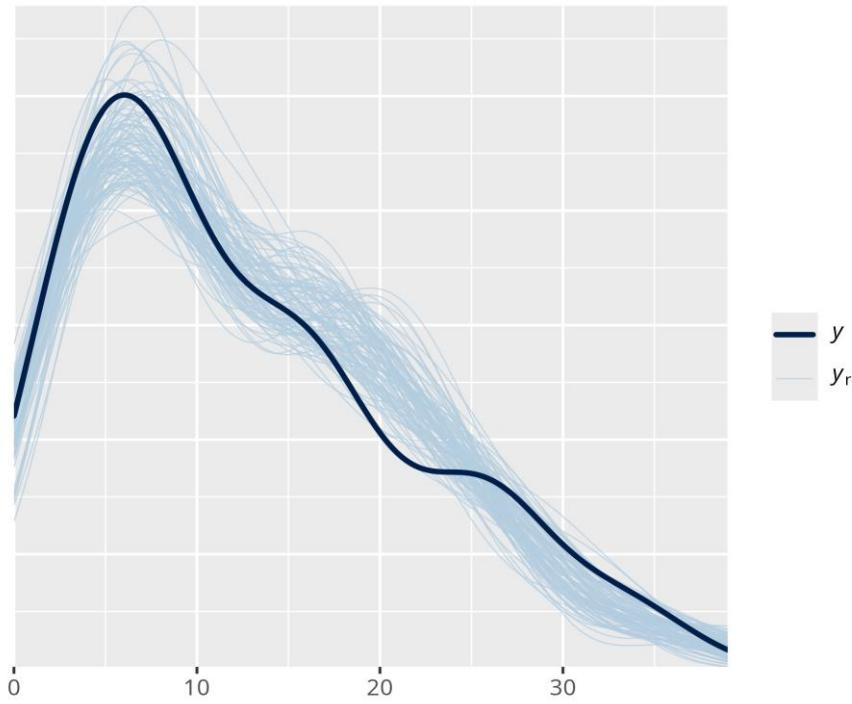


Response scale  $\mu$

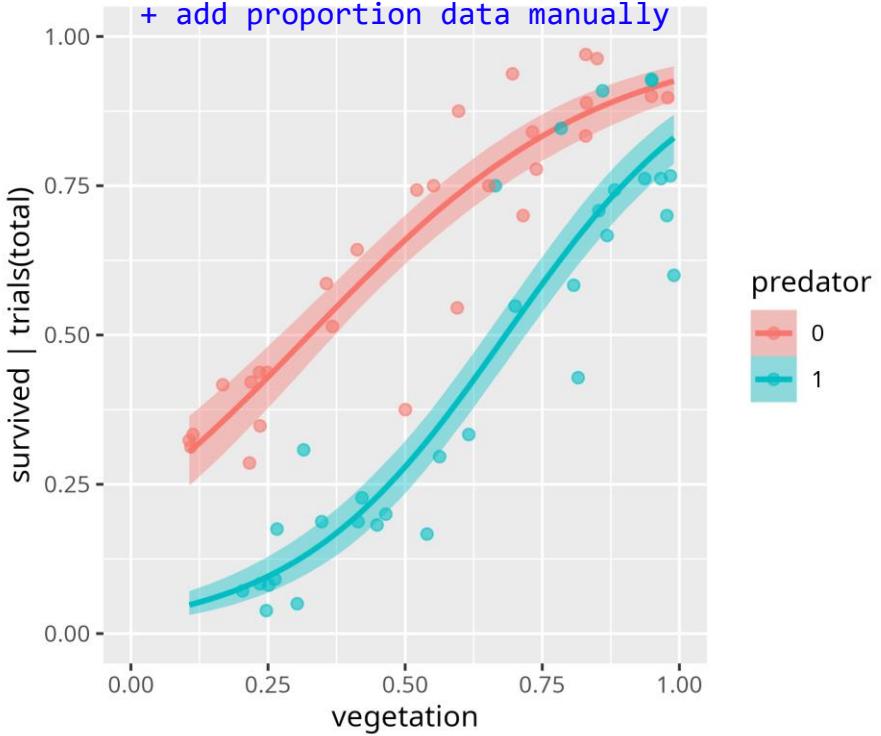


# Binomial regression: evaluation

```
> pp_check(fit, ndraws=100)
```

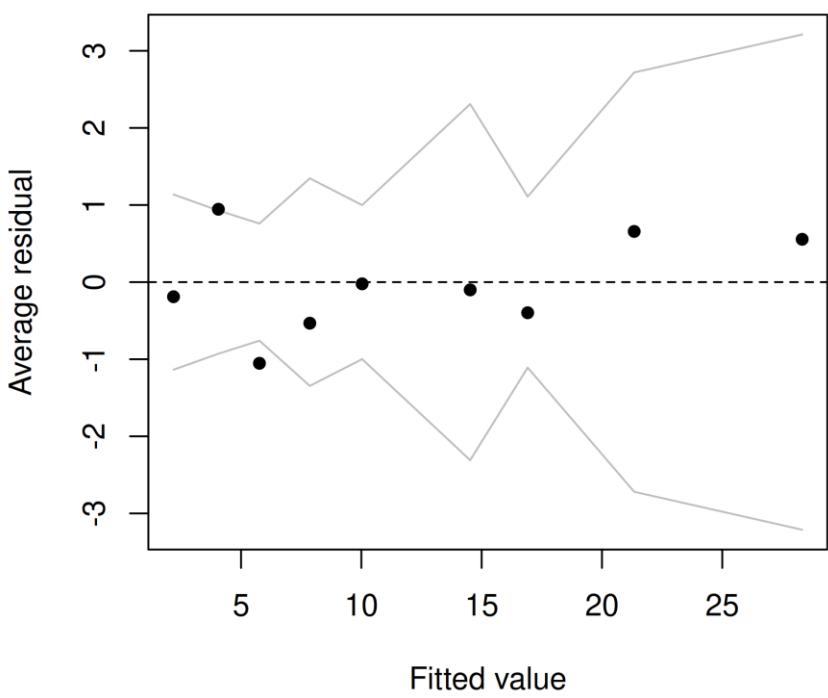


```
> plot(conditional_effects(fit,...))  
+ add proportion data manually
```

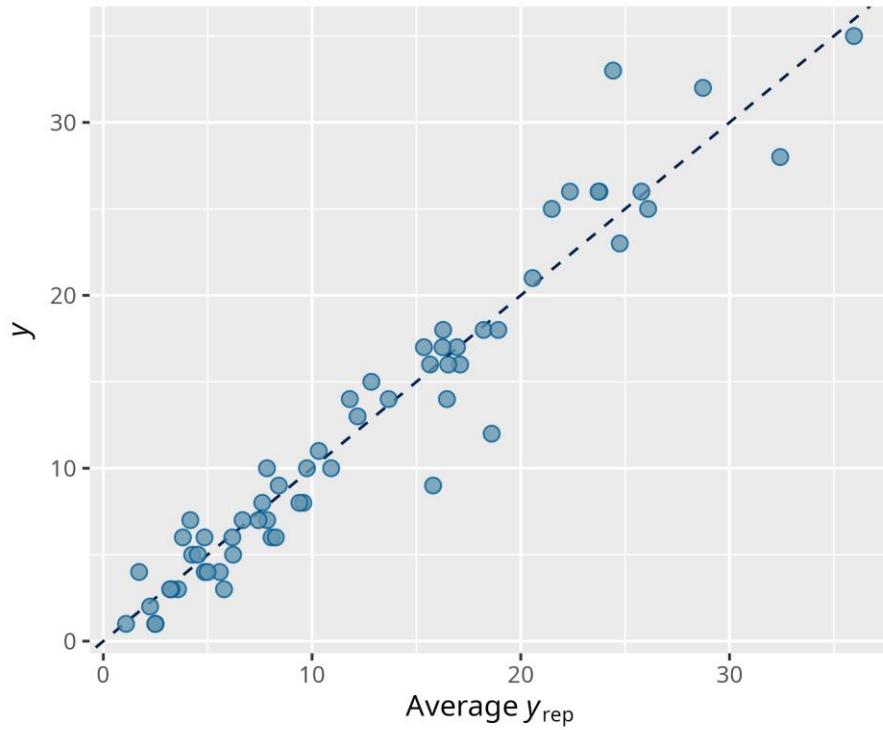


# Binomial regression: evaluation

```
> arm:::binnedplot(fitted, residuals)
```



```
> pp_check(fit, type=„scatter_avg“)
```



*Continuous example:  
Beta regression*

# Beta regression

**Example:** Deer survival probability  
versus vegetation and predation

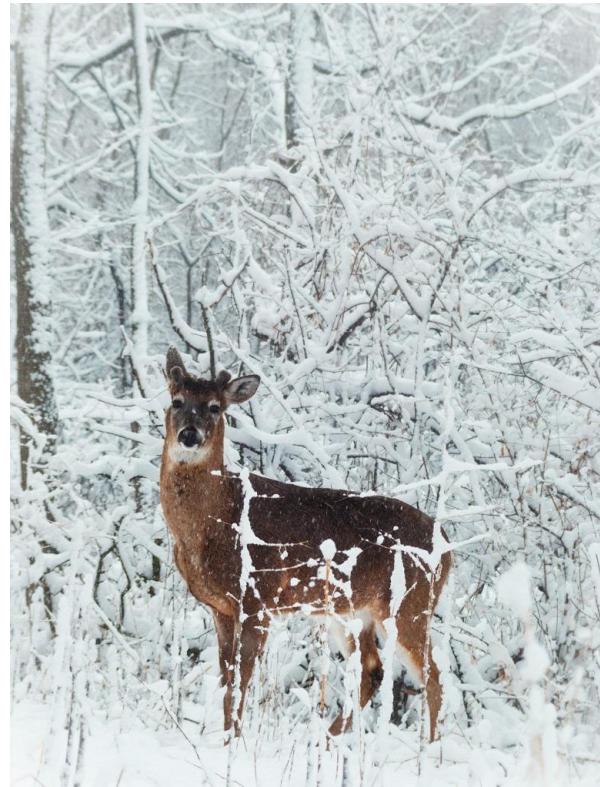
Same data as in Binomial regression.

But this time it's an old dataset where they only recorded  
the **ratio**  $\in [0, 1]$  of surviving individuals, not actual numbers

Bad practice: e.g. 50/100 contains more information than 1/2.

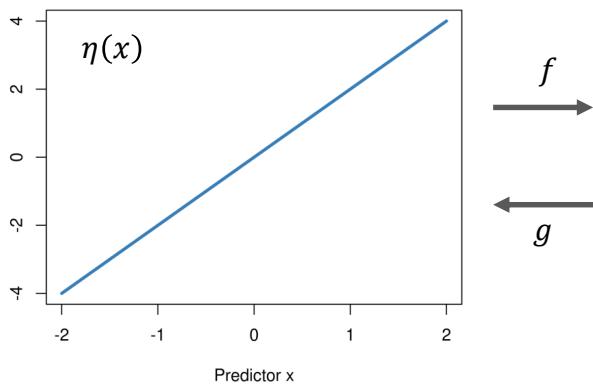
When transforming both to 0.5, this information is lost.

**Q:** How much does vegetation affect survival probability?  
Must control for predation.

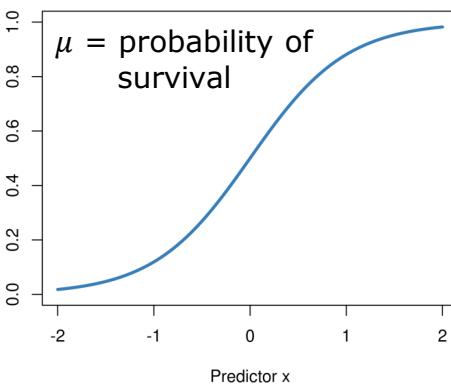


# Beta regression: overview

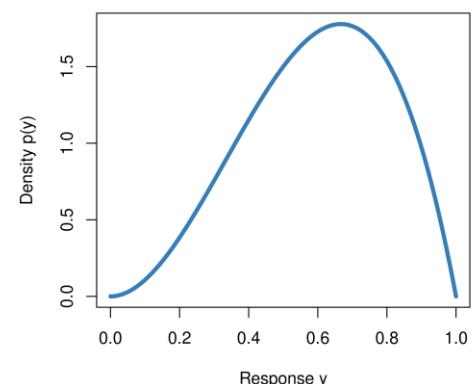
1: linear model



2: (inverse) link



3: residual distribution



**Logit link**

$$g = \text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$$

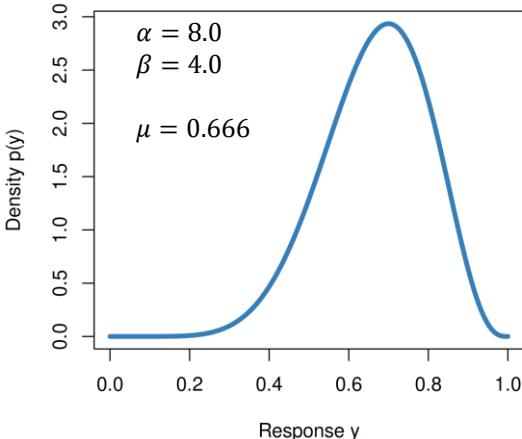
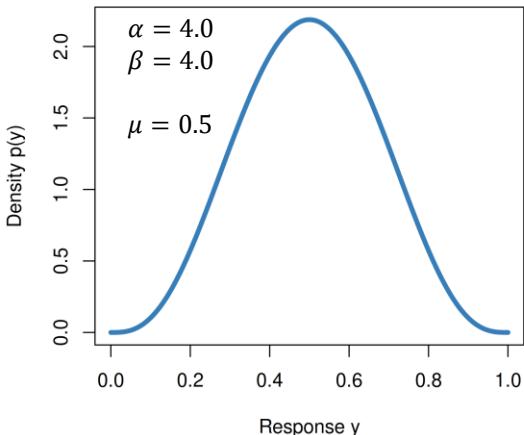
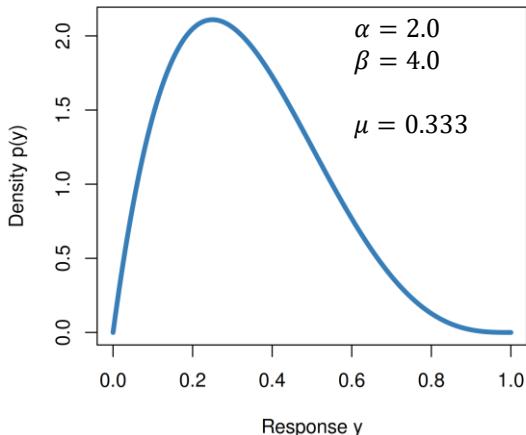
**Beta distribution**

$$y \sim \text{Beta}(\alpha, \beta)$$

Inverse-logit ("logistic")

$$f = \text{logistic}(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$$

# Beta distribution



Continuous distribution, bounded between 0 and 1.

Complicated parameterization with 2 shape parameters  $\alpha, \beta$ : Mean  $\frac{\alpha}{\alpha+\beta}$ . Sdev decreases with  $\alpha$  and  $\beta$

brms takes care of parameterization:

Mean  $\mu$  and shape parameter  $\phi$  generate  $\alpha, \beta$  automatically

ATTN: Exact  $y = 0$  or  $y = 1$  not allowed.  
→ Use zero- or one-inflated beta, or ordered beta.

# Beta regression: model fitting

```
> brm(proportion ~ vegetation * predator,
      family = Beta(link=logit),
      prior = ... )
```

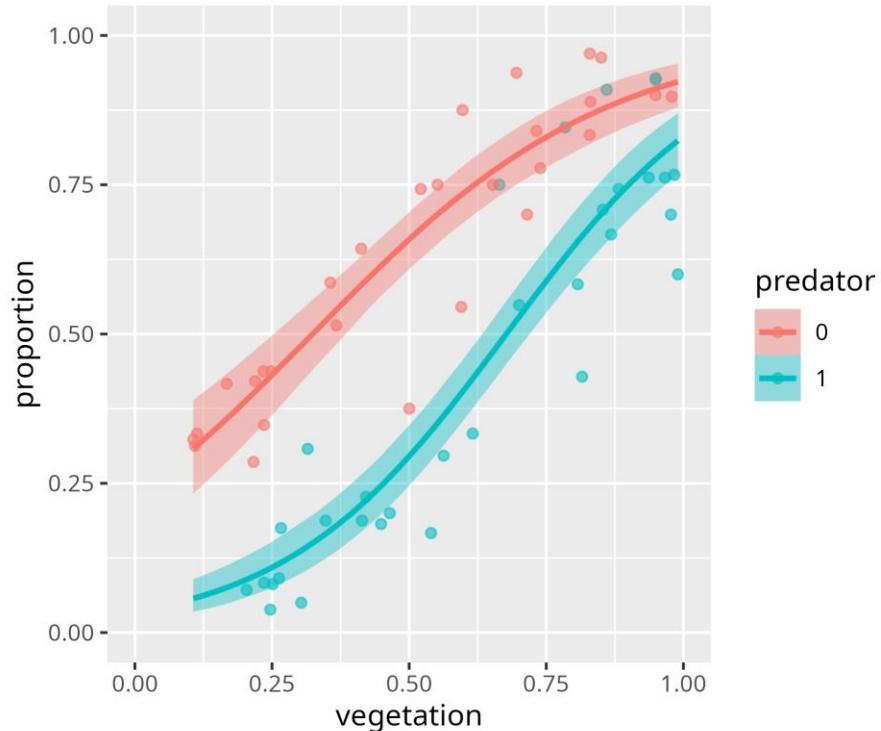
```
Family: beta
Links: mu = logit; phi = identity
Formula: proportion ~ scale(vegetation) * predator
Data: df (Number of observations: 60)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
      total post-warmup draws = 4000
```

```
Regression Coefficients:
Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Intercept          0.92     0.12     0.69    1.16 1.00   3467
scalevegetation    1.05     0.12     0.82    1.30 1.00   2425
predator1         -1.44     0.16    -1.76   -1.13 1.00   3824
scalevegetation:predator1 0.34     0.17    -0.01    0.67 1.00   2580
```

```
Further Distributional Parameters:
Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
phi      14.91     2.65    10.14    20.61 1.00   3450    2348
```

Results slightly different from Binomial regression

Additional scale parameter  $\phi \sim 1/\text{sdev}$



## *Poisson regression*

# Poisson regression

**Example:** Abundance of trout in stream transects.

Recorded number of individuals per transect.

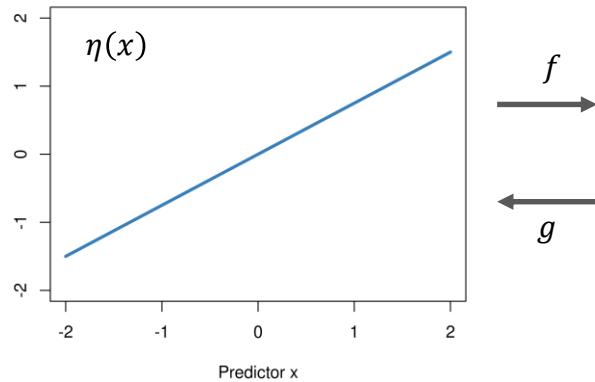
Measured temperature and concentration of a pollutant.

**Q:** Is pollutant concentration more harmful in warm streams?

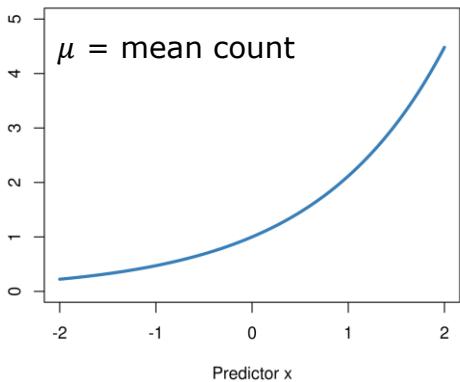


# Binomial regression: overview

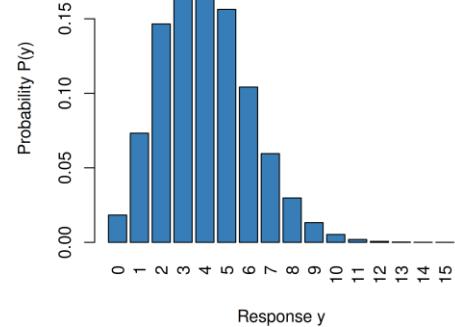
1: linear model



2: (inverse) link



3: residual distribution



**Log link**

$$g = \log(\mu)$$

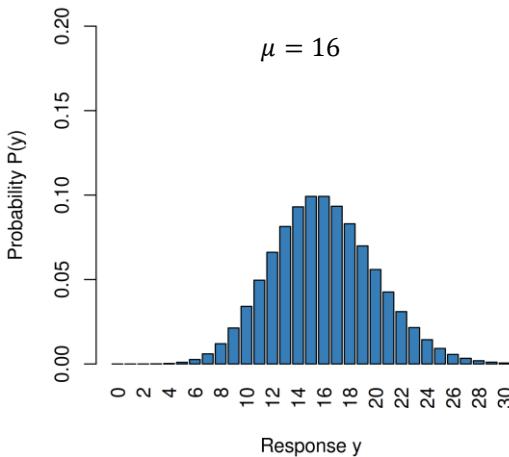
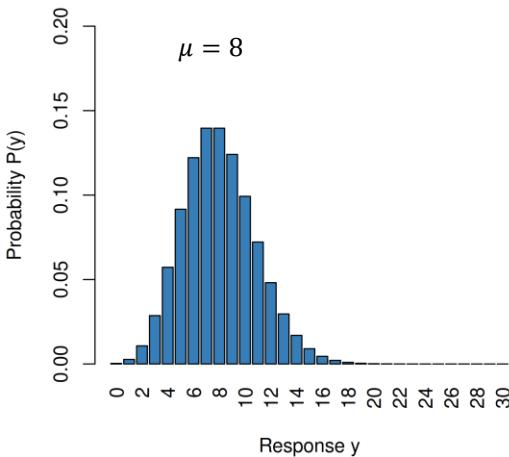
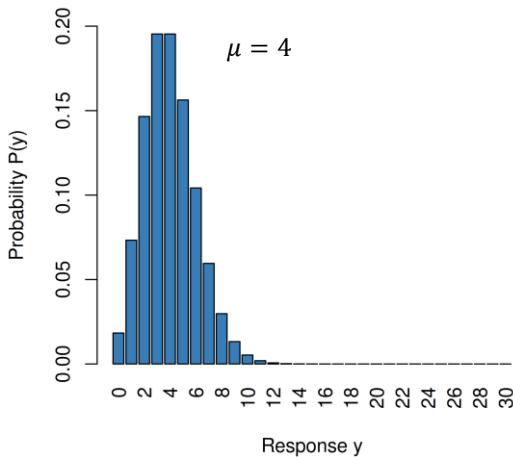
Inverse-link

$$f = \exp(\eta)$$

**Poisson distribution**

$$y \sim \text{Poisson}(\mu)$$

# Poisson distribution



Discrete distribution, lower bound 0, no upper boundary → used for counting data

Mean:  $\mu$

Sdev:  $\sqrt{\mu}$  (Variance equal to mean)

You make more errors counting many fish compared to when counting just a few fish

# Poisson regression: interaction model

## Deterministic part

(linear model & link)

$$\log(\mu) = b_0 + b_1 \cdot \text{temp} + b_2 \cdot \text{pollution}$$
$$+ b_3 \cdot \text{temp} \cdot \text{pollution}$$

## Stochastic part

$$\text{abundance} \sim \text{Poisson}(\mu)$$

## Priors

(scaled predictors!)

$$b_0 \sim \text{brms-default}$$

$$b_1 \sim \text{Normal}(0,1)$$

$$b_2 \sim \text{Normal}(-1,1) \text{ neg. pollution effect}$$

$$b_3 \sim \text{Normal}(0,1)$$

```
> brm(abundance ~ temperature * pollution,
       family = poisson(link=log),
       prior = ... )
```

	abundance	pollution	temperature
1	9	0.001	-0.418
2	12	-0.252	-1.650
3	17	-1.035	-1.266
4	2	2.749	-0.130
5	6	-0.377	1.982
6	7	0.318	2.096
7	12	-1.050	0.127
8	15	1.043	-1.989
9	16	-0.851	-0.295
10	14	-0.922	0.528
11	14	-0.503	-0.260
12	7	-0.740	0.437
13	6	1.391	-0.505
14	10	0.253	-0.980
15	14	0.263	-1.961

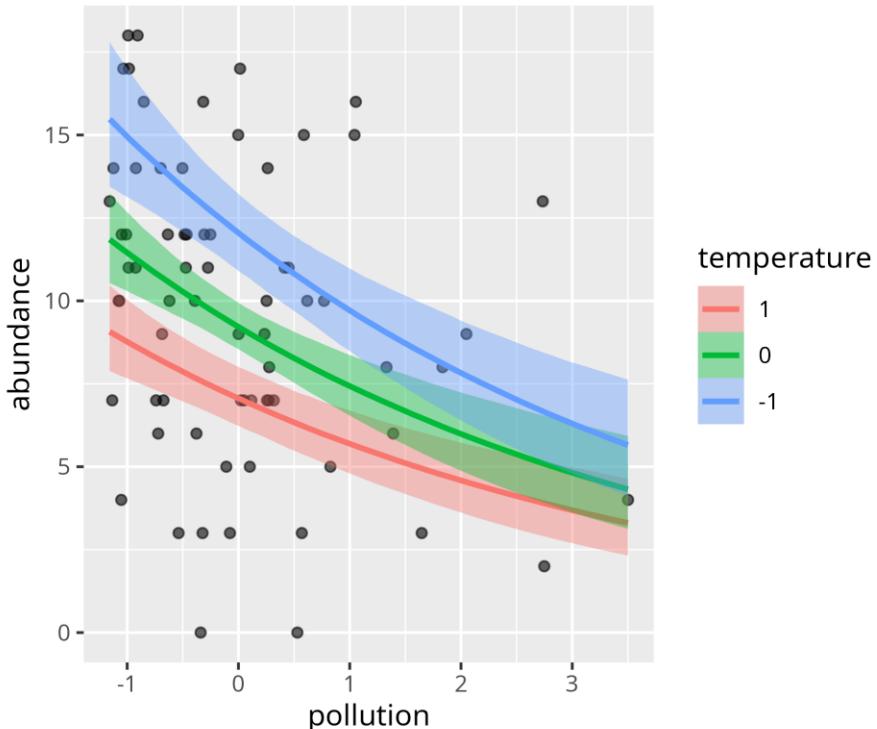
# Poisson regression: additive model

```
> brm(abundance ~ temperature + pollution,  
      family = poisson(link=log),  
      prior = ... )
```

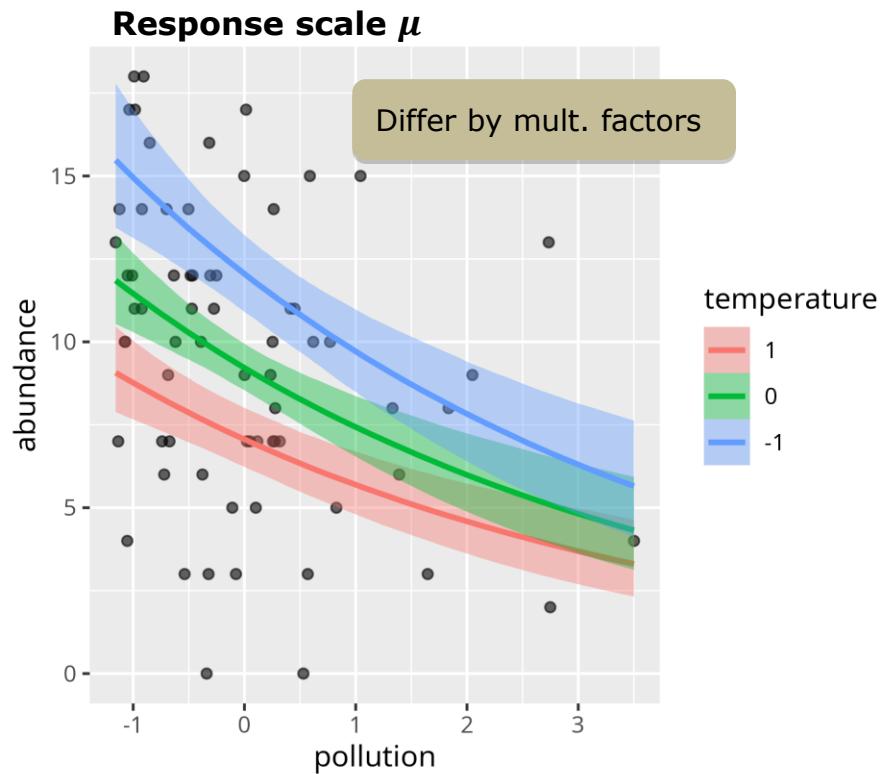
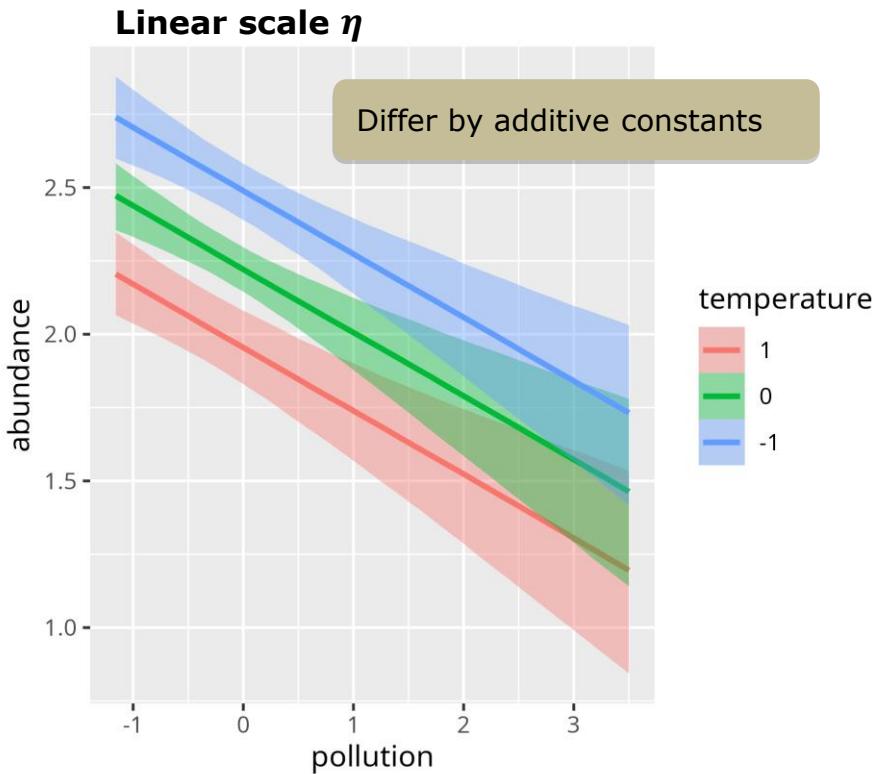
```
Family: poisson  
Links: mu = log  
Formula: abundance ~ temperature + pollution  
Data: df (Number of observations: 70)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
       total post-warmup draws = 4000  
  
Regression Coefficients:  
Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS  
Intercept     2.22      0.04    2.15    2.30 1.00  2807  
temperature   -0.27      0.04   -0.35   -0.19 1.00  3102  
pollution    -0.22      0.04   -0.30   -0.13 1.00  3512
```

→ Temperature and pollution both have negative effects on abundance

(slopes are on linear scale)



# Poisson regression: additive model



# Poisson regression: interaction model

```
> brm(abundance ~ temperature * pollution,
      family = poisson(link=log),
      prior = ... )
```

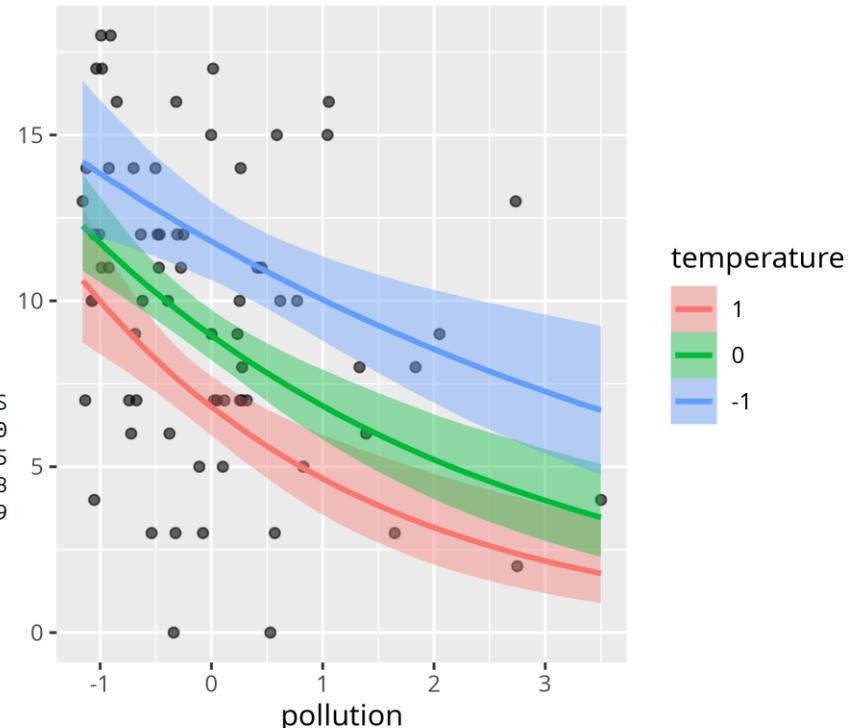
```
Family: poisson
Links: mu = log
Formula: abundance ~ temperature * pollution
Data: df (Number of observations: 70)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
      total post-warmup draws = 4000
```

Regression Coefficients:

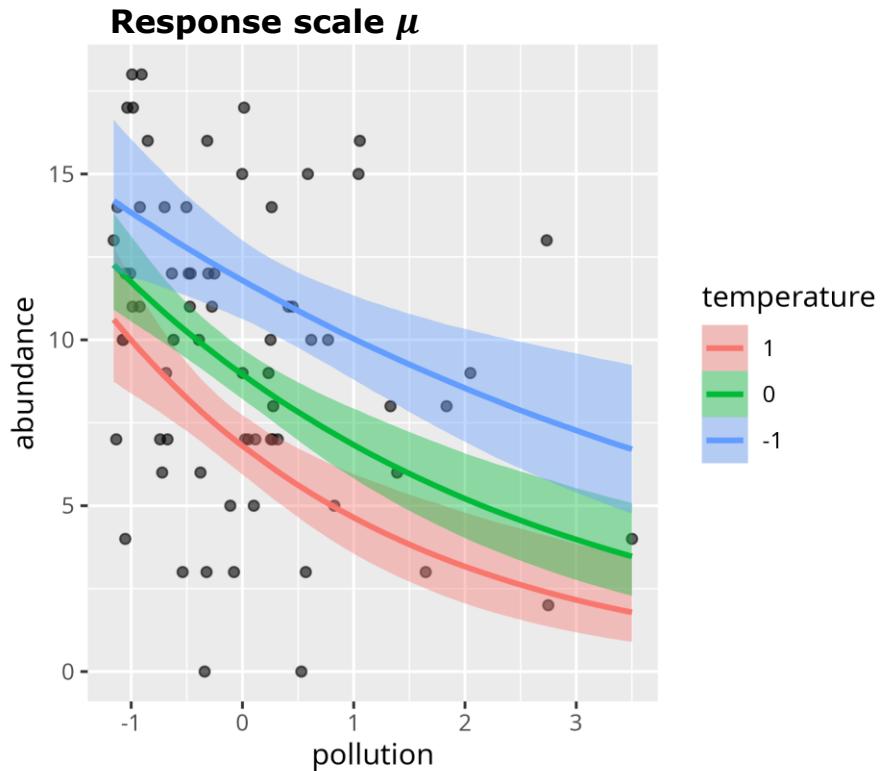
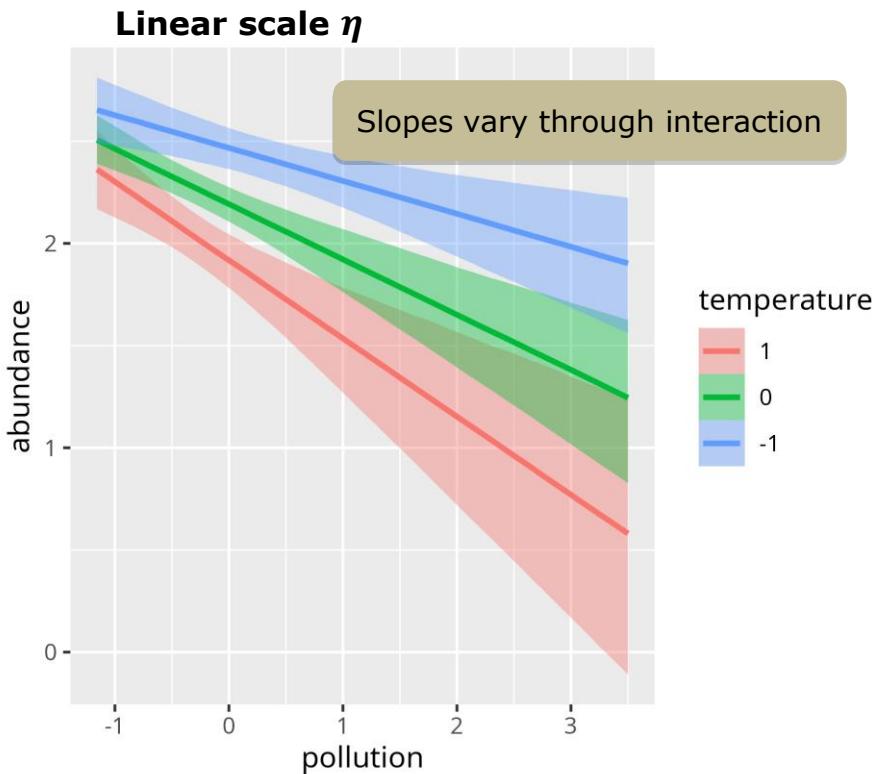
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS
Intercept	2.19	0.04	2.11	2.27	1.00	3490
temperature	-0.27	0.04	-0.36	-0.20	1.00	3545
pollution	-0.27	0.05	-0.38	-0.17	1.00	3378
temperature:pollution	-0.11	0.05	-0.21	-0.01	1.00	3419

**Q:** Is pollutant concentration more harmful in warm streams?

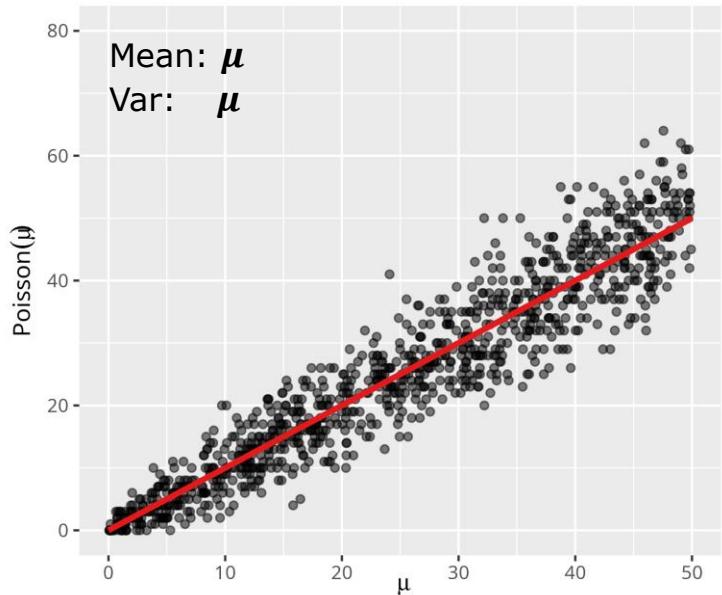
→ Yes, negative pollution effect gets stronger with temp.



# Poisson regression: interaction model

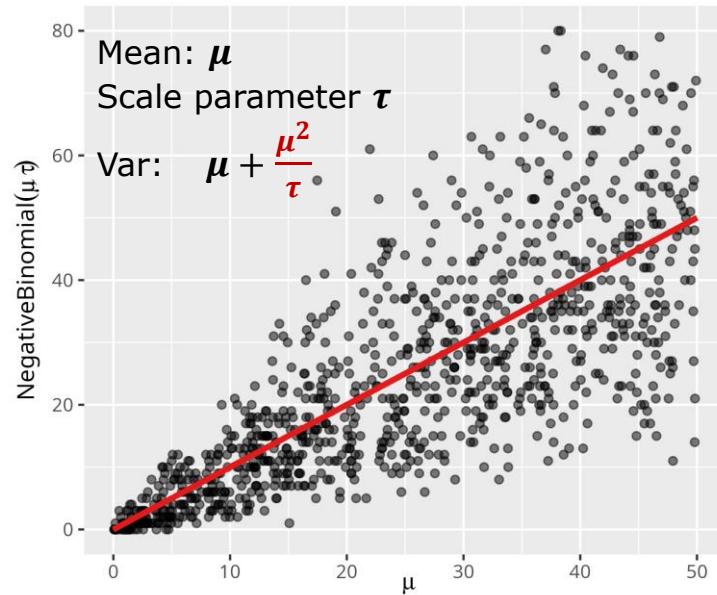


# Poisson regression: Overdispersion



Often count data does not follow the mean-variance relation of Poisson

→ Use Negative Binomial distribution



$\tau$  small  $\rightarrow \text{Var} \sim \mu^2$   
 $\tau$  big  $\rightarrow \text{Var} \sim \mu$

> `family=negbinomial(link=log)`

*Continuous example:  
Distributional model*

# Distributional models

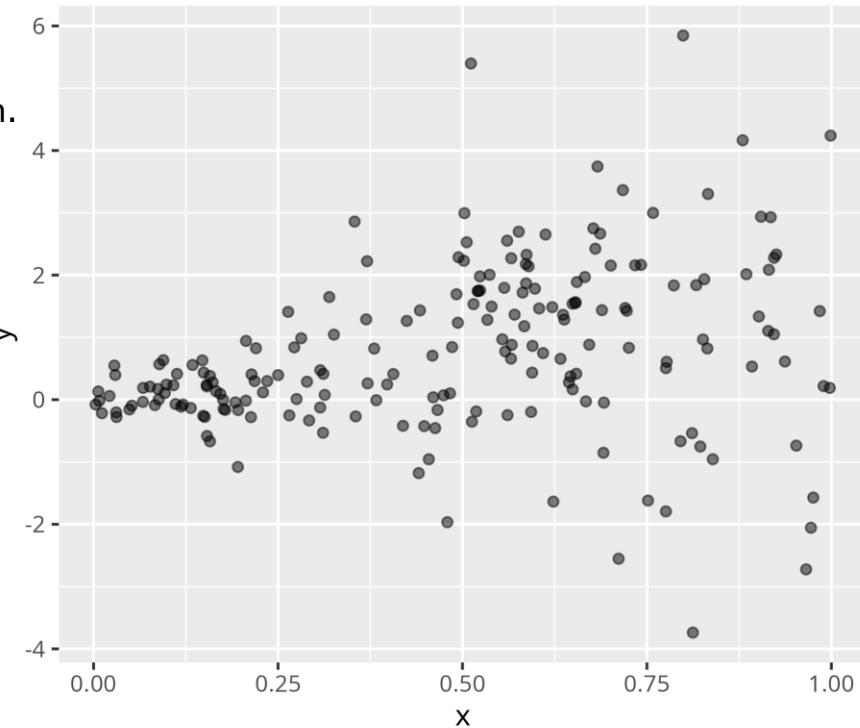
Data is clearly **heteroskedastic** (non-constant sdev)

Continuous response  $y \rightarrow$  Can't use Poisson / Neg.Bin.

Linear regression line wanted  $\rightarrow$  Can't use log-link

We can make **standard deviation dependent on  $x$**

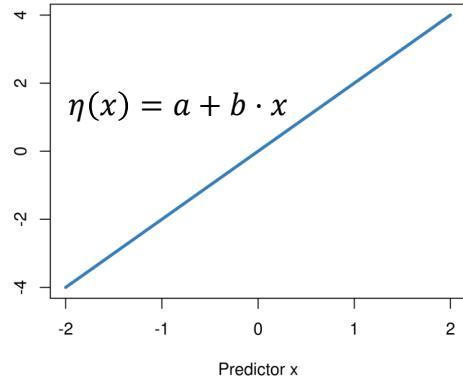
$\rightarrow y \sim \text{Normal}(\mu(x), \sigma(x))$



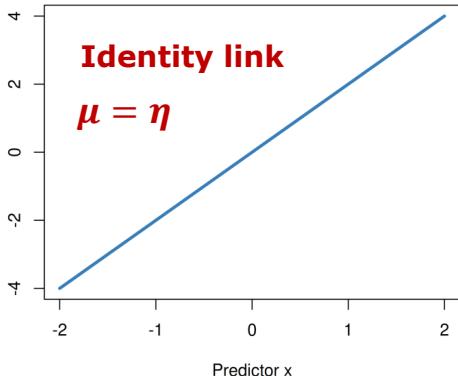
# Distributional models

## Mean model

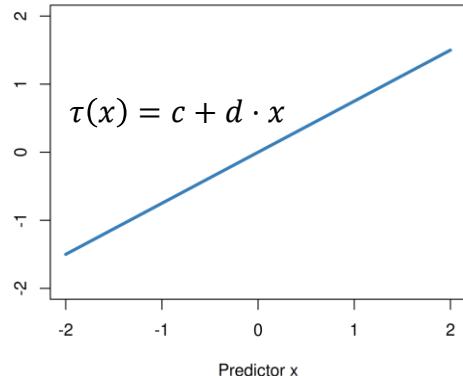
1: linear model



2: (inverse) link

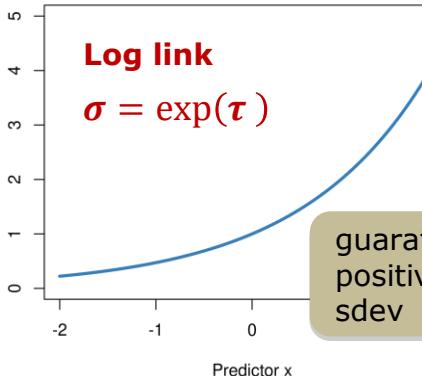


## Sdev model

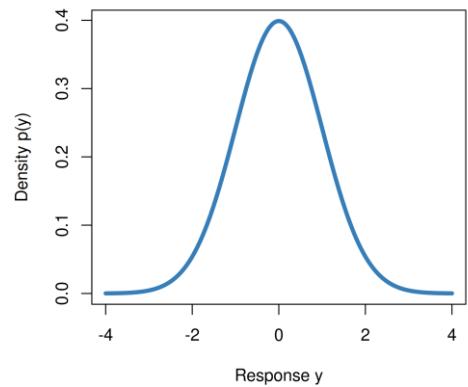


Log link  
 $\sigma = \exp(\tau)$

guarantees positive sdev



3: residual distribution



$$y \sim \text{Normal}(\mu(x), \sigma(x))$$

# Regular model fit

```
> fit_lm = brm( y ~ x )
```

```
Family: gaussian  
Links: mu = identity; sigma = identity  
Formula: y ~ x  
Data: df (Number of observations: 200)  
Draws: 4 chains, each with iter = 2000; warmup  
total post-warmup draws = 4000
```

## Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.09	0.18	-0.25	0.43
x	1.39	0.32	0.75	2.02

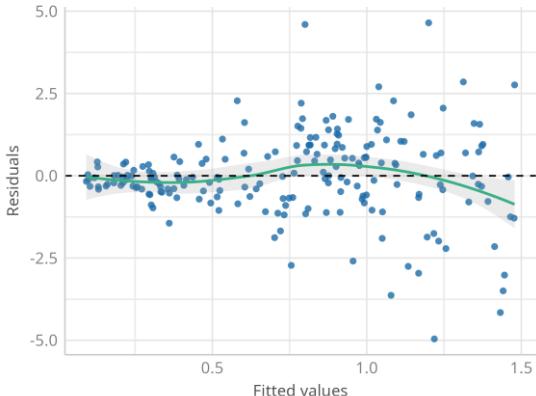
## Further Distributional Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
sigma	1.29	0.07	1.17	1.42	1.00

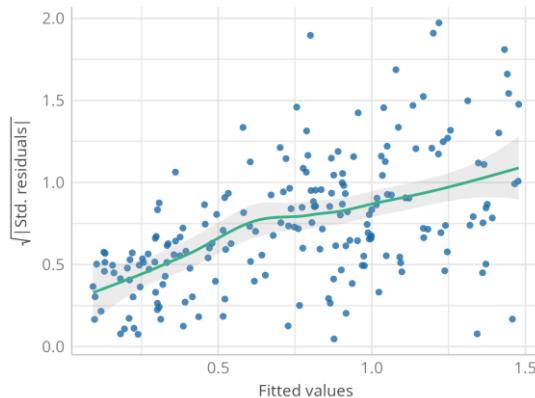
Bad model fit !

Some assumptions violated

Linearity  
Reference line should be flat and horizontal



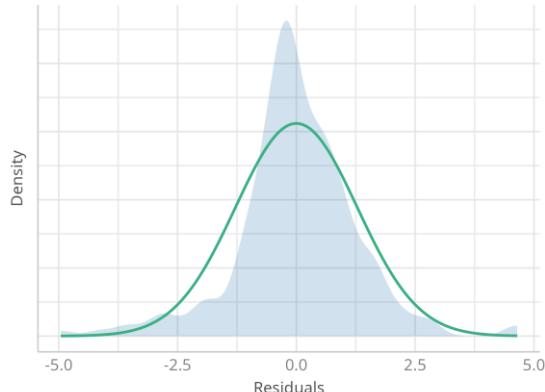
Homogeneity of Variance  
Reference line should be flat and horizontal



Normality of Residuals  
Dots should fall along the line



Normality of Residuals  
Distribution should be close to the normal curve



# Distributional model fit

```
> fit_distr = brm( bf( y~x, sigma~x ),  
family = gaussian() )
```

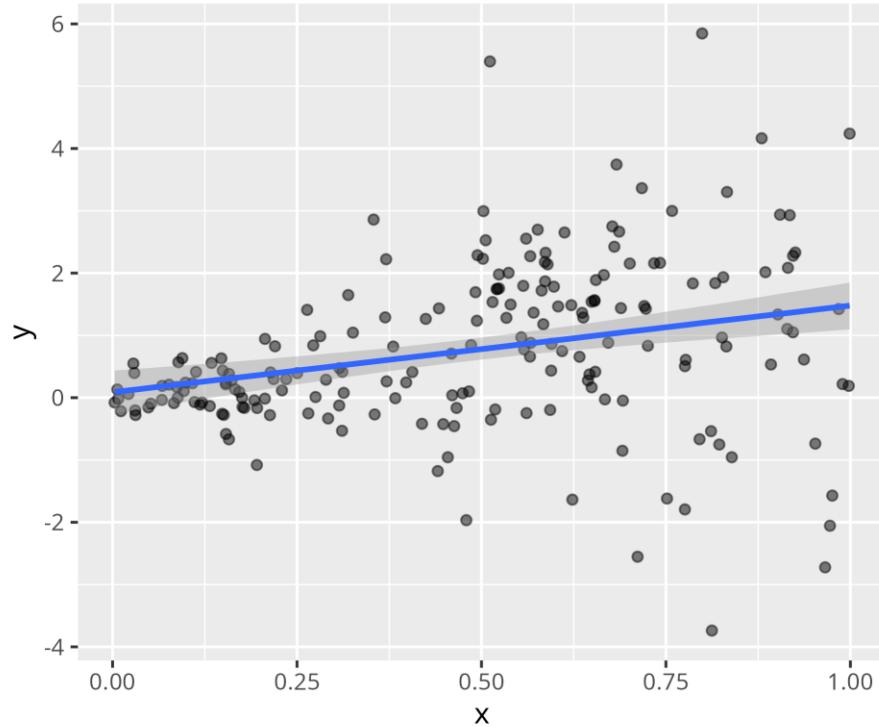
```
Family: gaussian  
Links: mu = identity; sigma = log  
Formula: y ~ x  
sigma ~ x  
Data: df (Number of observations: 200)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
total post-warmup draws = 4000
```

## Regression Coefficients:

	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat	Bulk_ESS
Intercept	-0.09	0.07	-0.23		0.05	1.00		3776
sigma_Intercept	-1.12	0.11	-1.34		-0.89	1.00		3416
x	1.87	0.25	1.37		2.37	1.00		1987
sigma_x	2.29	0.21	1.88		2.72	1.00		3375

Log link for sigma as default

Slope in x substantially different from `fit_lm`

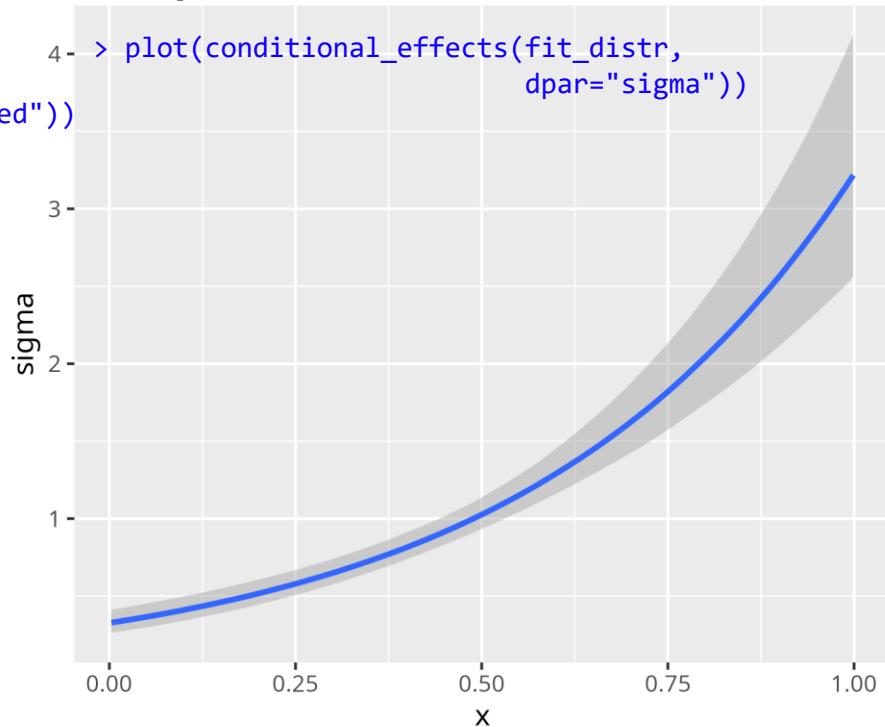


# Distributional model fit

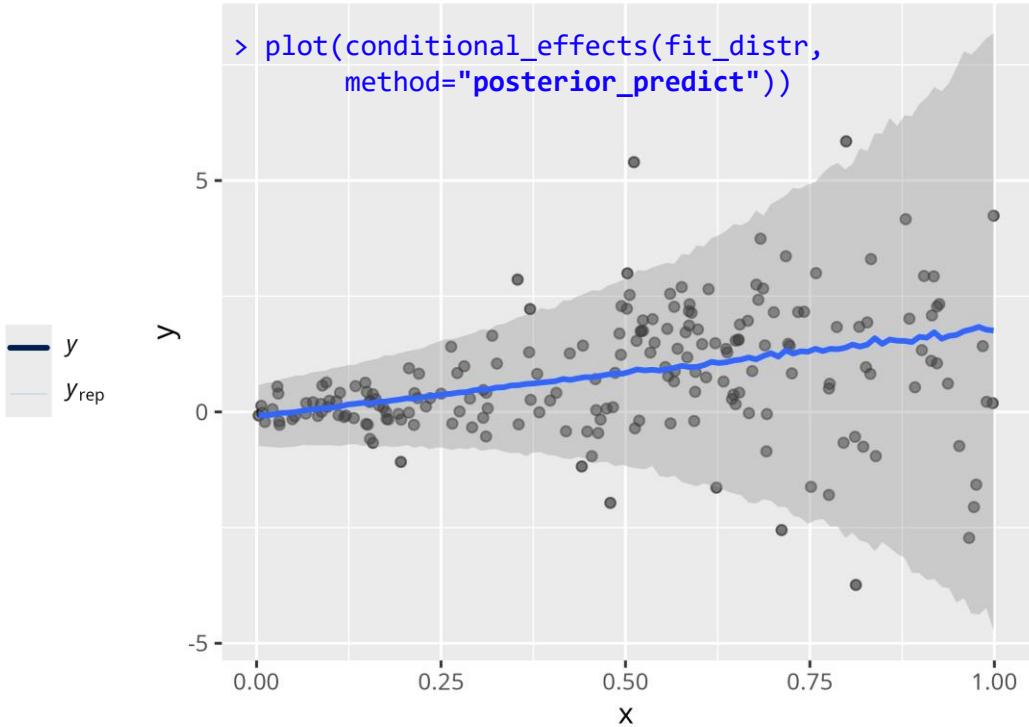
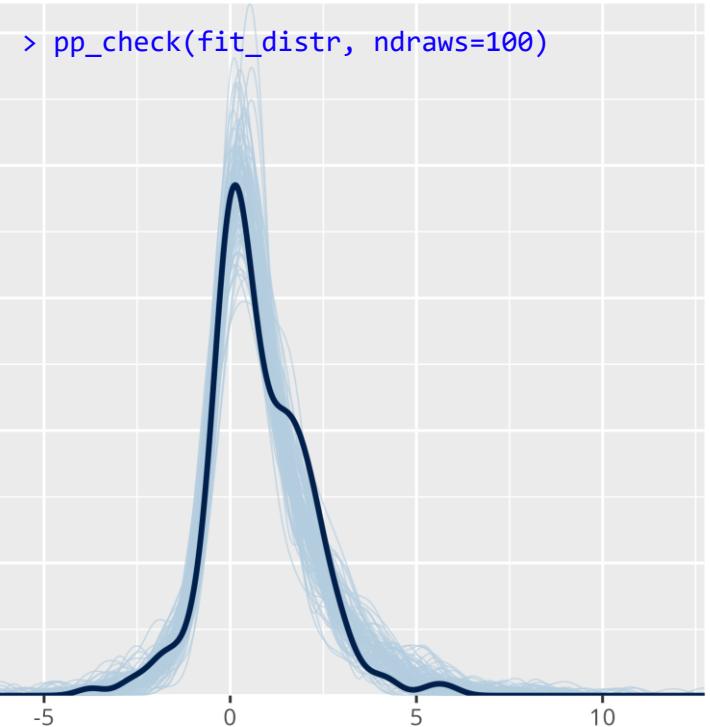
Linear scale  $\tau$



Response scale  $\sigma$



# Distributional model: evaluation



# Distributional models: ANOVA

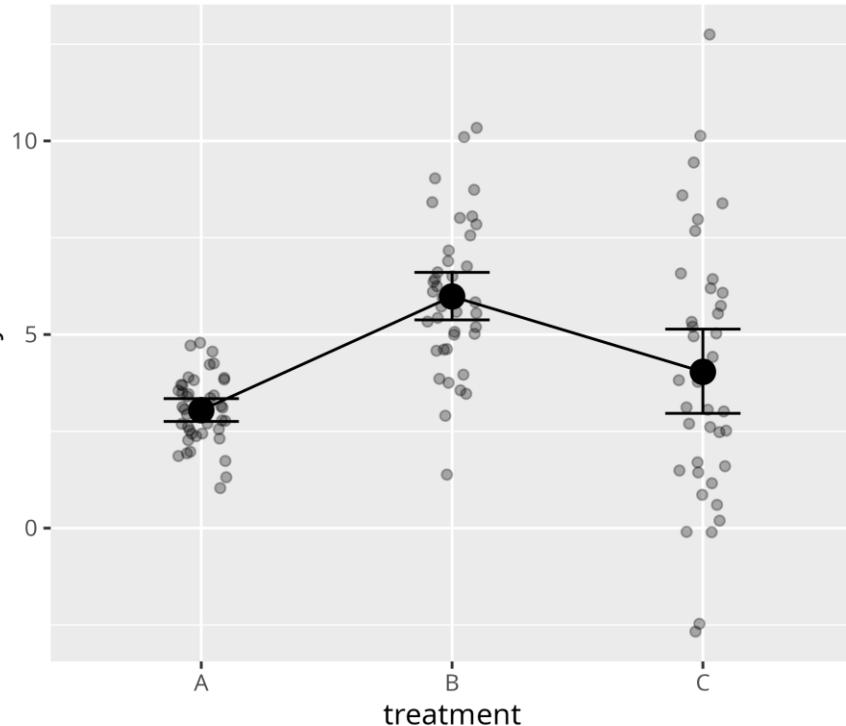
```
> fit.anova = brm( bf( y ~ 0+treatment,
                      sigma ~ 0+treatment ),
                      family = gaussian() )
```

Family: gaussian  
Links: mu = identity; sigma = log  
Formula: y ~ 0 + treatment  
sigma ~ 0 + treatment  
Data: df (Number of observations: 120)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1; > 5  
total post-warmup draws = 4000

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS
treatmentA	3.05	0.15	2.76	3.34	1.00	5711
treatmentB	5.99	0.31	5.38	6.61	1.01	5946
treatmentC	4.04	0.55	2.96	5.14	1.00	5347
sigma_treatmentA	-0.09	0.11	-0.31	0.14	1.00	5740
sigma_treatmentB	0.67	0.11	0.45	0.89	1.00	6186
sigma_treatmentC	1.23	0.11	1.02	1.46	1.00	6352

on linear scale.  
→  $sdev = \exp(\sigma)$



## *Beyond GLMs: Nonlinear models*

# Nonlinear model

**Example:** Microcosm experiment

Feeding rates of a small crustacean

"Functional response"  $F(N) = \frac{aN}{1+ahN}$

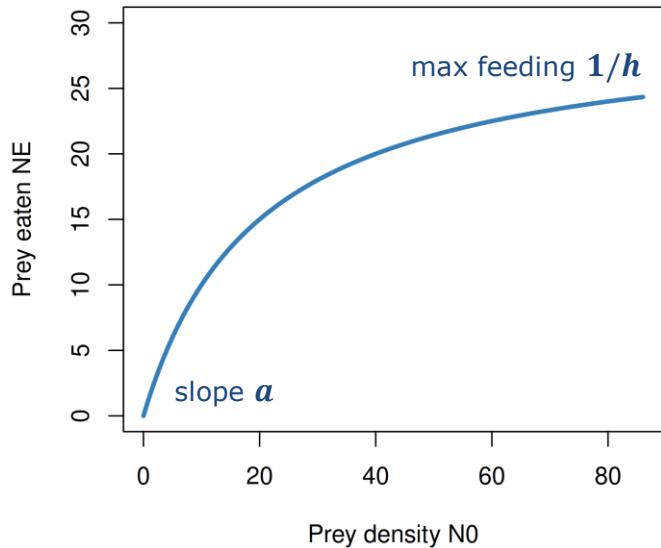
Depends on prey density  $N$

Want to estimate **attack rate  $a$**  and **handling time  $h$**

**Data:**

Controlled feeding trials in the lab (1 hour)

at different prey densities  $N_0$ .



# Nonlinear model: fitting

```
FR.formula = bf( NE ~ a*N0/(1+a*h*N0),      → the model formula
                  a ~ 1,                      → parameters do not depend
                  h ~ 1,                      on other predictors
                  nl = TRUE)                 → it's a nonlinear model

FR.priors  = c(prior(exponential(1.0), nlpar="a", lb=0),   → exp. distr. mean=1, sdev=1
               prior(exponential(1.0), nlpar="h", lb=0))       and positive (lower bound 0)

fit.joint = brm(FR.formula,
                 prior  = FR.priors,
                 family = poisson(link=identity), → response=counts, no log-link required
                 data   = df)                         (FR model is always positive)
```

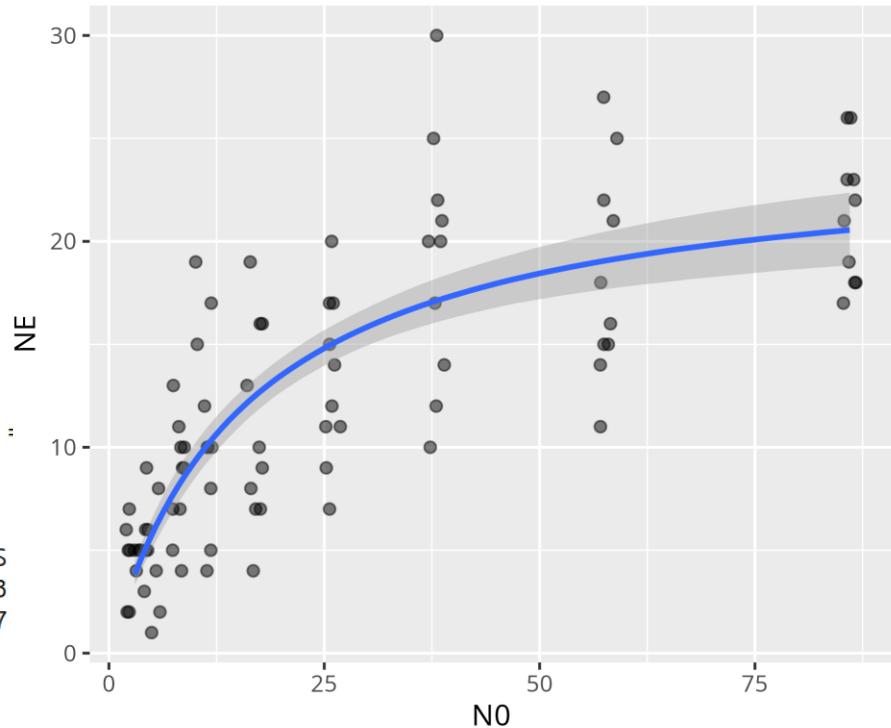
# Nonlinear model: fitting

```
FR.formula = bf( NE ~ a*N0/(1+a*h*N0),  
                 a ~ 1,  
                 h ~ 1,  
                 nl = TRUE)
```

```
Family: poisson  
Links: mu = identity  
Formula: NE ~ a * N0/(1 + a * h * N0)  
        a ~ 1  
        h ~ 1  
Data: df (Number of observations: 90)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin =  
       total post-warmup draws = 4000
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS
a_Intercept	1.52	0.15	1.26	1.85	1.01	1403
h_Intercept	0.04	0.00	0.04	0.05	1.00	1467



# Nonlinear model: fitting

Include **categorical predictor** temperature (levels = low / high)

**Q:** Is feeding behavior different in warm experiments?

```
FR.formula = bf( NE ~ a*N0/(1+a*h*N0),      → the model formula  
                 a ~ 0 + temp,                → individual parameters at low / high temp.  
                 h ~ 0 + temp,                0 to avoid dummy-coding  
                 nl = TRUE)                  → it's a nonlinear model
```

```
FR.priors = c(prior(exponential(1.0), nlpar="a", lb=0),   → exp. distr. mean=1, sdev=1  
              prior(exponential(1.0), nlpar="h", lb=0))       and positive (lower bound 0)
```

```
FR.fit.tmp = brm(FR.formula,  
                  prior = FR.priors,  
                  family = poisson(link=identity), → response=counts, no log-link required  
                  data   = df)                                (FR model is always positive)
```

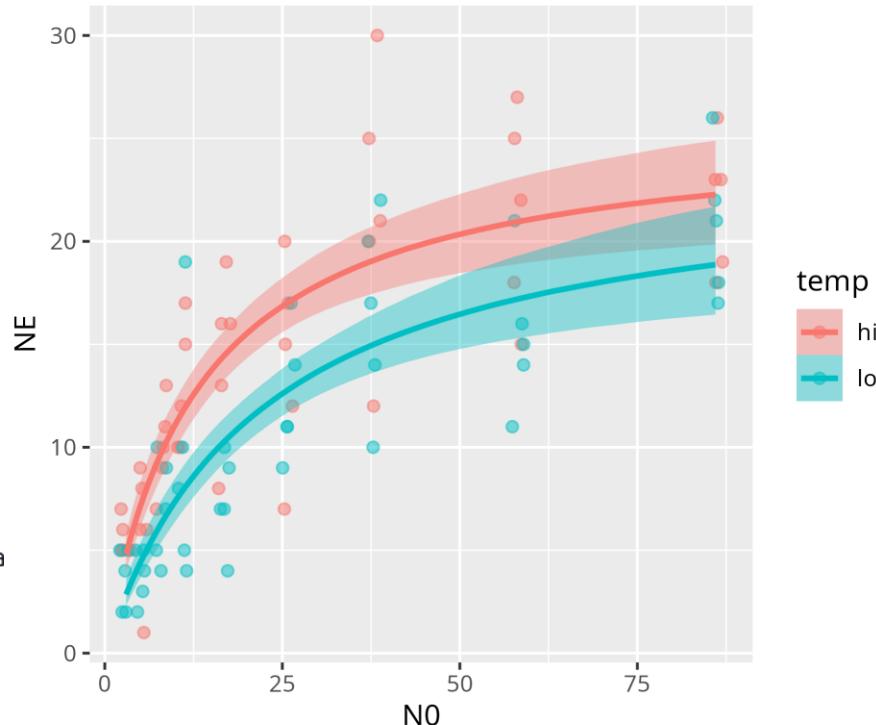
# Nonlinear model: fitting

```
FR.formula = bf( NE ~ a*N0/(1+a*h*N0),  
                 a ~ 0 + temp,  
                 h ~ 0 + temp,  
                 nl = TRUE)
```

```
Family: poisson  
Links: mu = identity  
Formula: NE ~ a * N0/(1 + a * h * N0)  
        a ~ 0 + temp  
        h ~ 0 + temp  
Data: df (Number of observations: 90)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin  
       total post-warmup draws = 4000
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Ta
a_temp_hi	2.00	0.26	1.56	2.57	1.00	2515	
a_temp_lo	1.10	0.15	0.84	1.44	1.00	2277	
h_temp_hi	0.04	0.00	0.03	0.05	1.00	2663	
h_temp_lo	0.04	0.00	0.03	0.05	1.00	2381	



# Nonlinear model: fitting

```
> hypothesis(FR.fit.tmp,
  c("a_temphi > a_templo",
    "h_temphi > h_templo" ))
```

Hypothesis Tests for class b:

	Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Evid.Ratio	Post.Prob	Star
1	(a_temphi)-(a_templo) > 0	0.9	0.30	0.43	1.42	1332.33	1.00	*
2	(h_temphi)-(h_templo) > 0	0.0	0.01	-0.01	0.01	0.41	0.29	
---								

'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

'\*': For one-sided hypotheses, the posterior probability exceeds 95%;  
for two-sided hypotheses, the value tested against lies outside the 95%-CI.  
Posterior probabilities of point hypotheses assume equal prior probabilities.

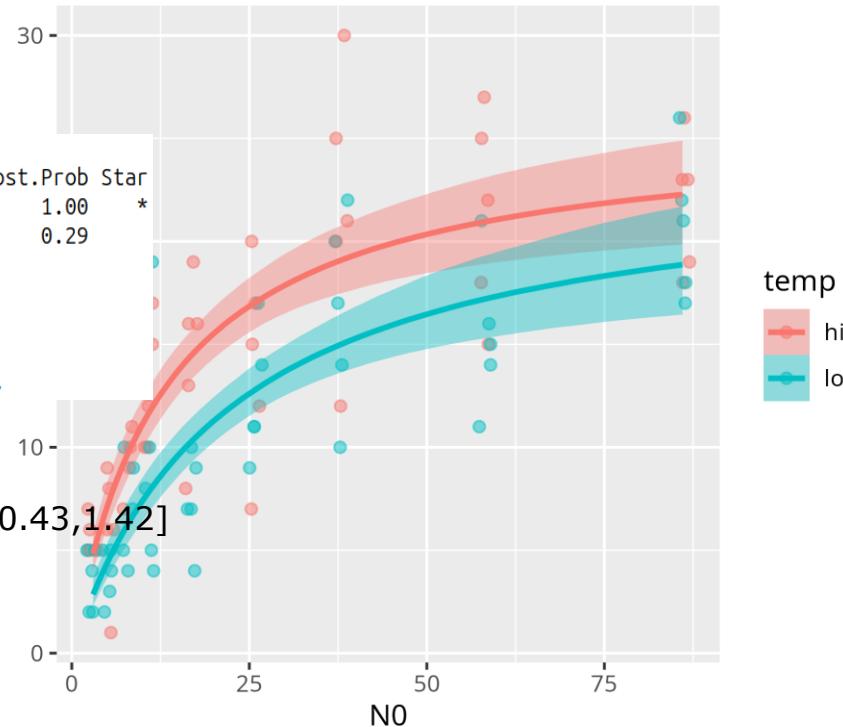
**Q:** Is feeding behavior different in warm experiments?

- Yes: „warm“ attack rates higher, mean diff. = 0.9 [0.43, 1.42]  
But no difference in handling time (max feeding)

```
> LOO(FR.fit.tmp, FR.fit)
```

Model comparisons:

	elpd_diff	se_diff
FR.fit.tmp	0.0	0.0
FR.fit	-11.0	5.6



# What's Bayesian about it?

- Exact quantification of parameter uncertainty.  
Even more important when moving beyond simple LMs.
- Don't use parameters' point estimates for prediction
  - Use full posterior predictive distribution
  - emmeans for contrasts
- Works well even for very small datasets
- Not limited to GLM framework
- No extra packages needed, do it all in brms

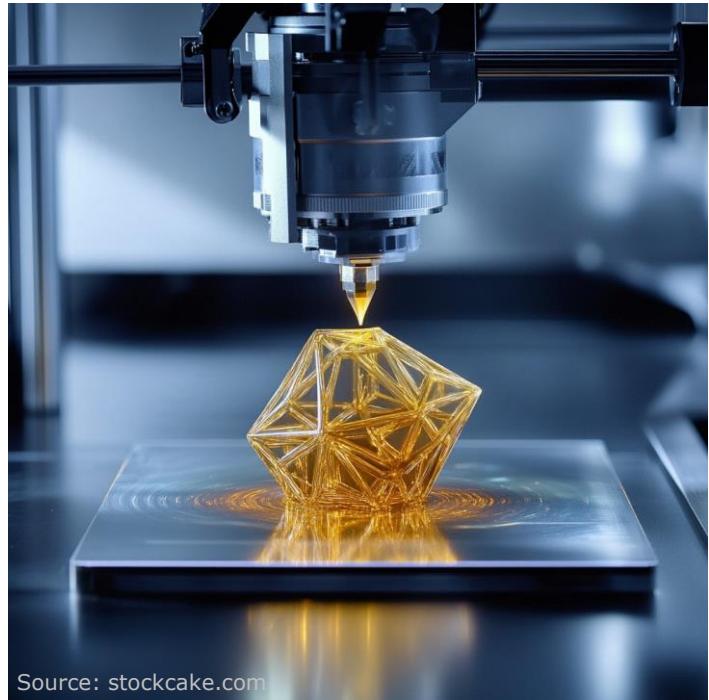


Source: Wikipedia

# Summary

- brms highly flexible, can fit vast amount of GLMs.  
Even nonlinear models!
- Bayesian stats not limited to GLM,  
but linear part in GLM is useful and interpretable
- Data should indicate which GLM (distribution & link) to use
- Be careful with priors when using link functions.  
Scaled / mean-centered predictors are your friend
- Use appropriate plots for model evaluation
- ATTN: Don't do model comparison (loo)  
for discrete vs. continuous residual distribution  
E.g. Poisson vs. Gaussian

The Bayesian 3D printer



Source: stockcake.com

# Further reading

Buckley, Y. M. (2015). Generalized linear models. In Fox, G. A. , Negrete-Yankelevich, S. and Sosa, V. J. (eds). Ecological Statistics: Contemporary theory and application. *Oxford Academic*.  
<https://doi.org/10.1093/acprof:oso/9780199672547.003.0007>

Bürkner, P. (2024). The brms Book [in progress]. <https://paulbuerkner.com/software/brms-book/> [Chapter 3]

Fieberg, J. (2024). Statistics 4 Ecologists. <https://statistics4ecologists-v2.netlify.app/> [Chapters 14-17]

Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and Other Stories. *Cambridge University Press*.  
<https://doi.org/10.1017/9781139161879> [Part 3]

Warton, D. (2022). Eco-Stats: Data Analysis in Ecology. *Springer (Methods in Statistical Ecology)* [Chapter 10]

Warton, D., Lyons, M., Stoklosa, J., & Ives, A. R. (2016). Three points to consider when choosing a LM or GLM test for count data. *Methods in Ecology and Evolution*, 7(8), 882–890. <https://doi.org/10.1111/2041-210X.12552>