# Intro to linear models and generalized linear models

Bodo Winter

UNIVERSITY OF BIRMINGHAM

@BodoWinter

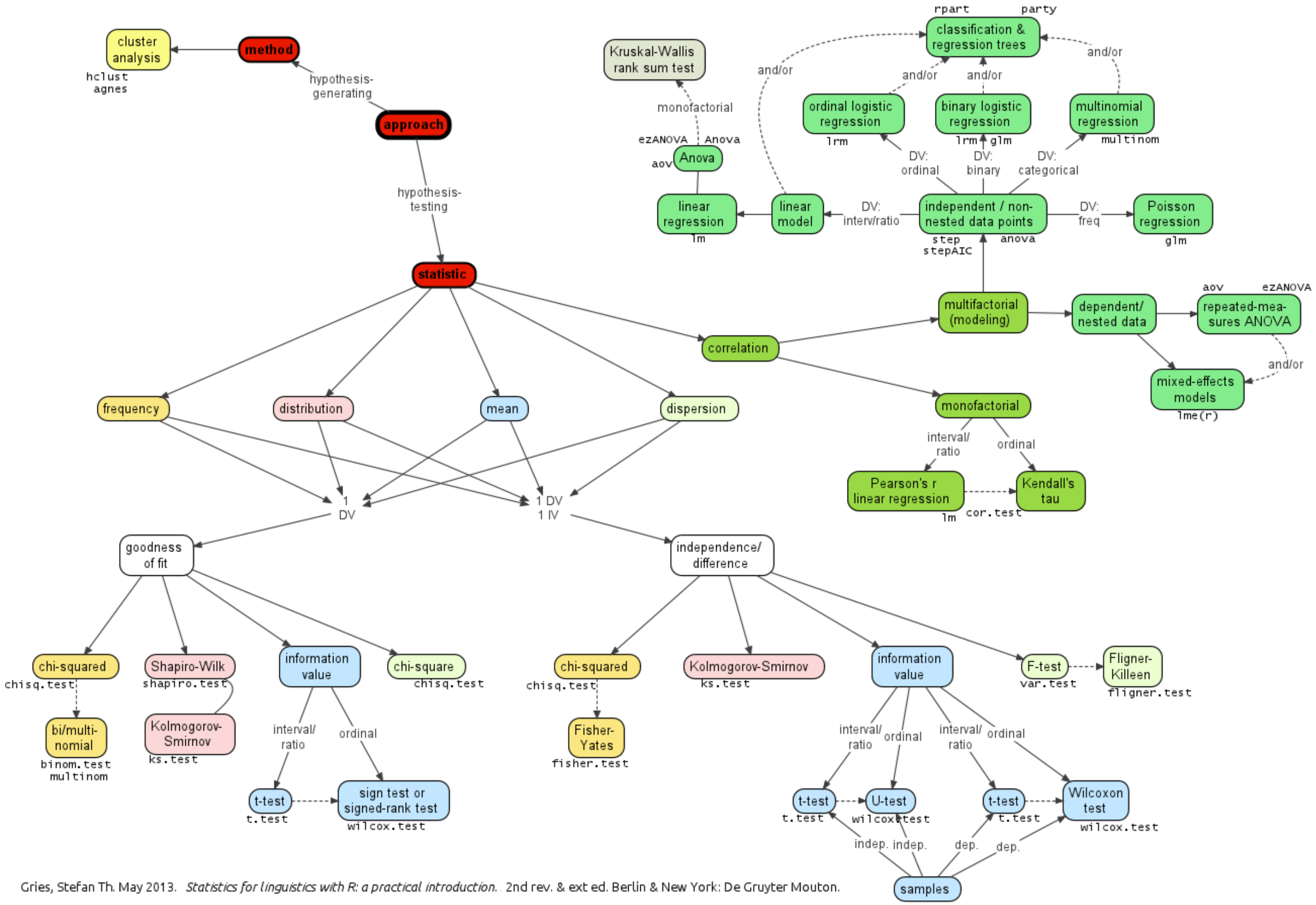
www.bodowinter.com

Exercise with `broom`

Linear models 101

Intro to GLM

# The messy world of statistical testing




Gries, Stefan Th. May 2013. *Statistics for linguistics with R: a practical introduction.* 2nd rev. & ext ed. Berlin & New York: De Gruyter Mouton.

# Linear models

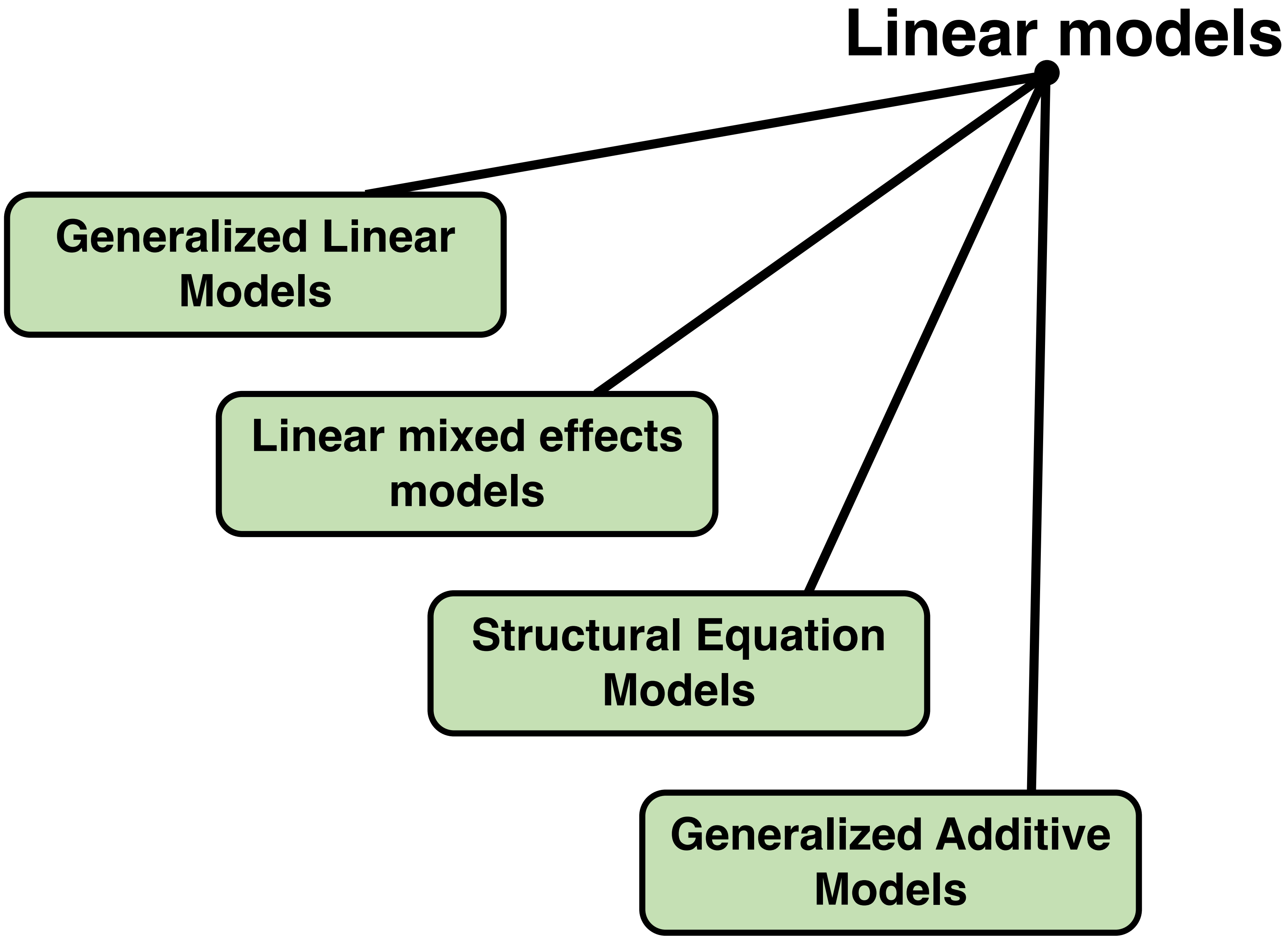
Simple, well-understood

Interpretable

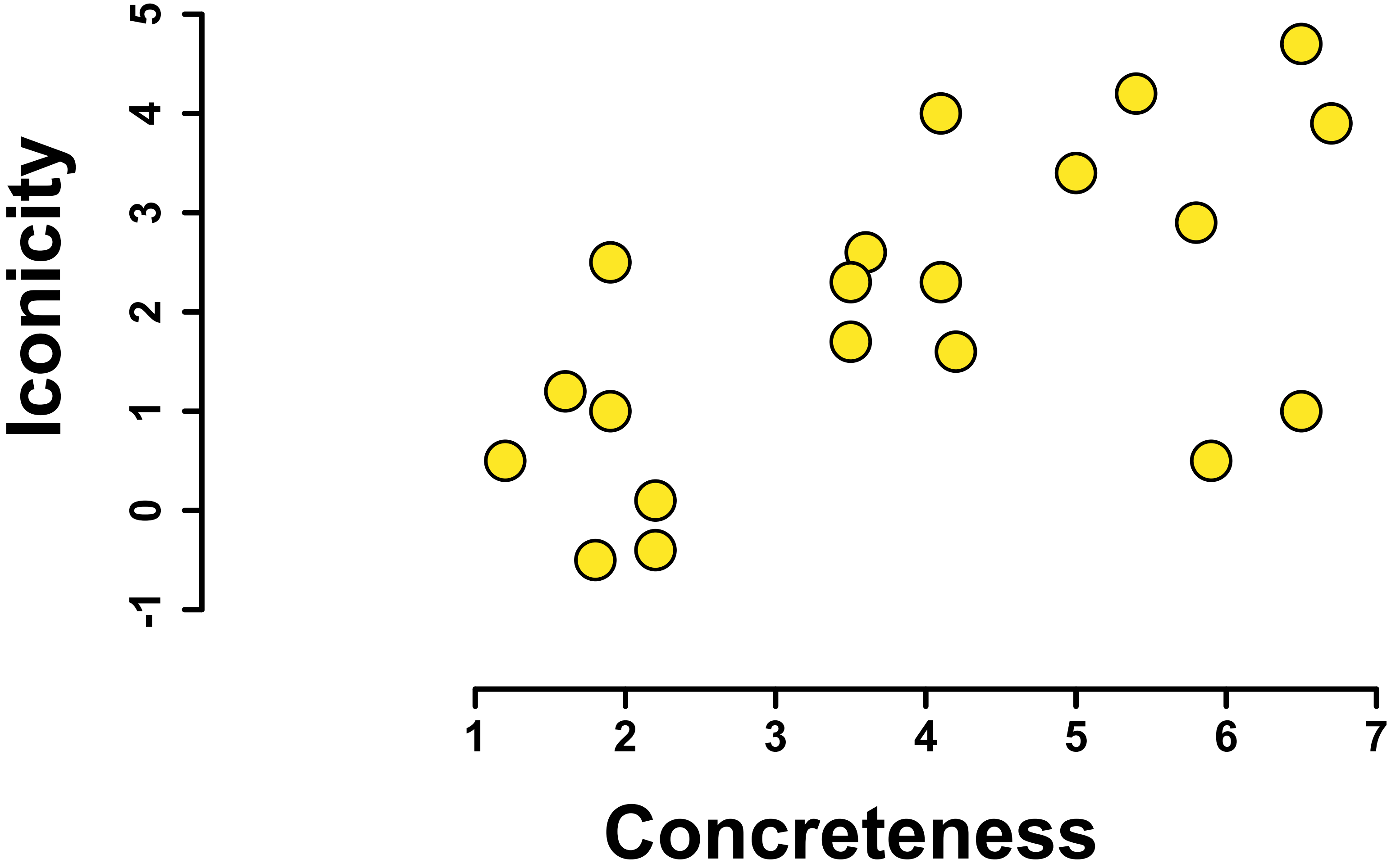
Good for theory-driven work

Flexible

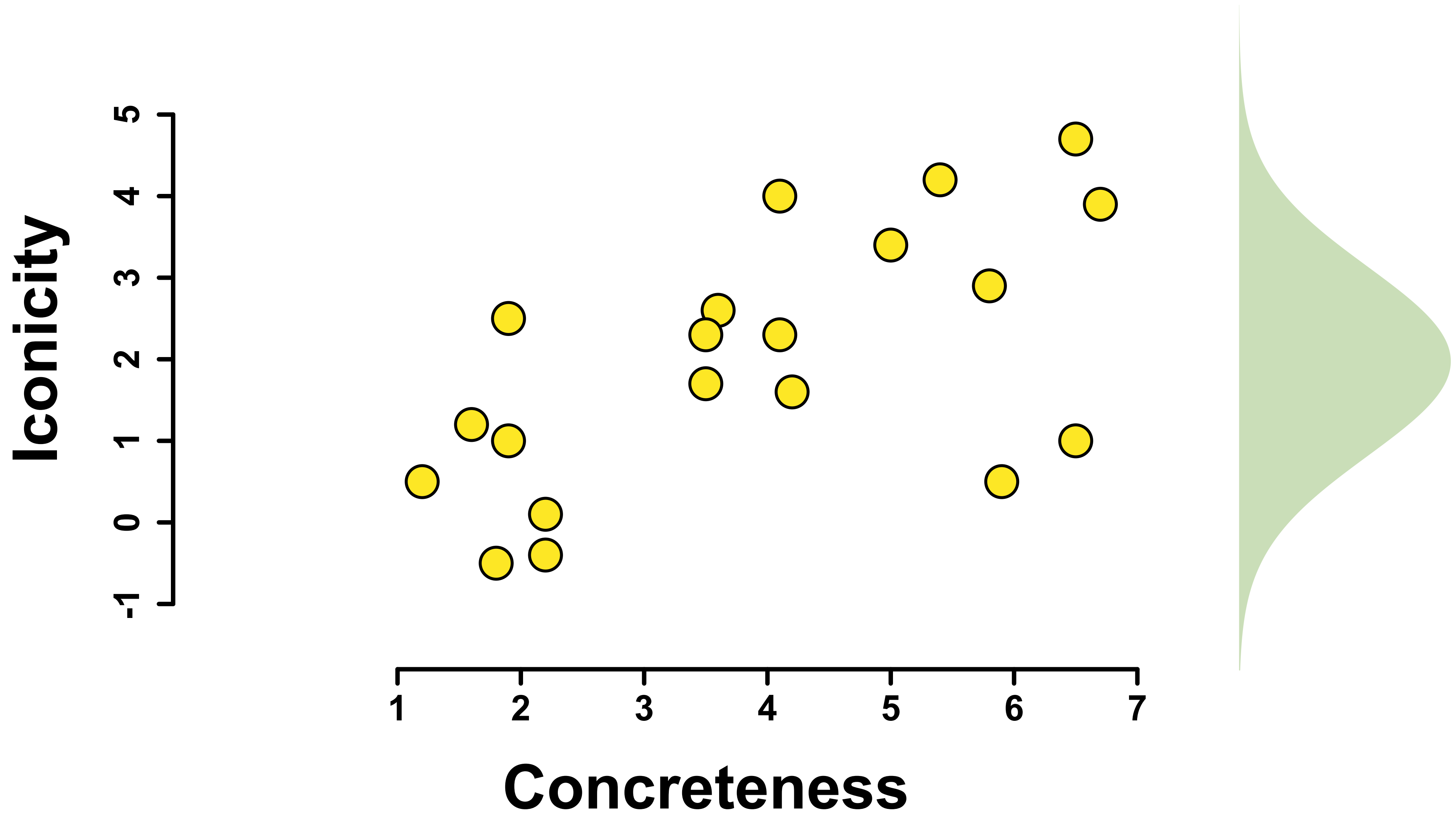
→ Bayesian analysis

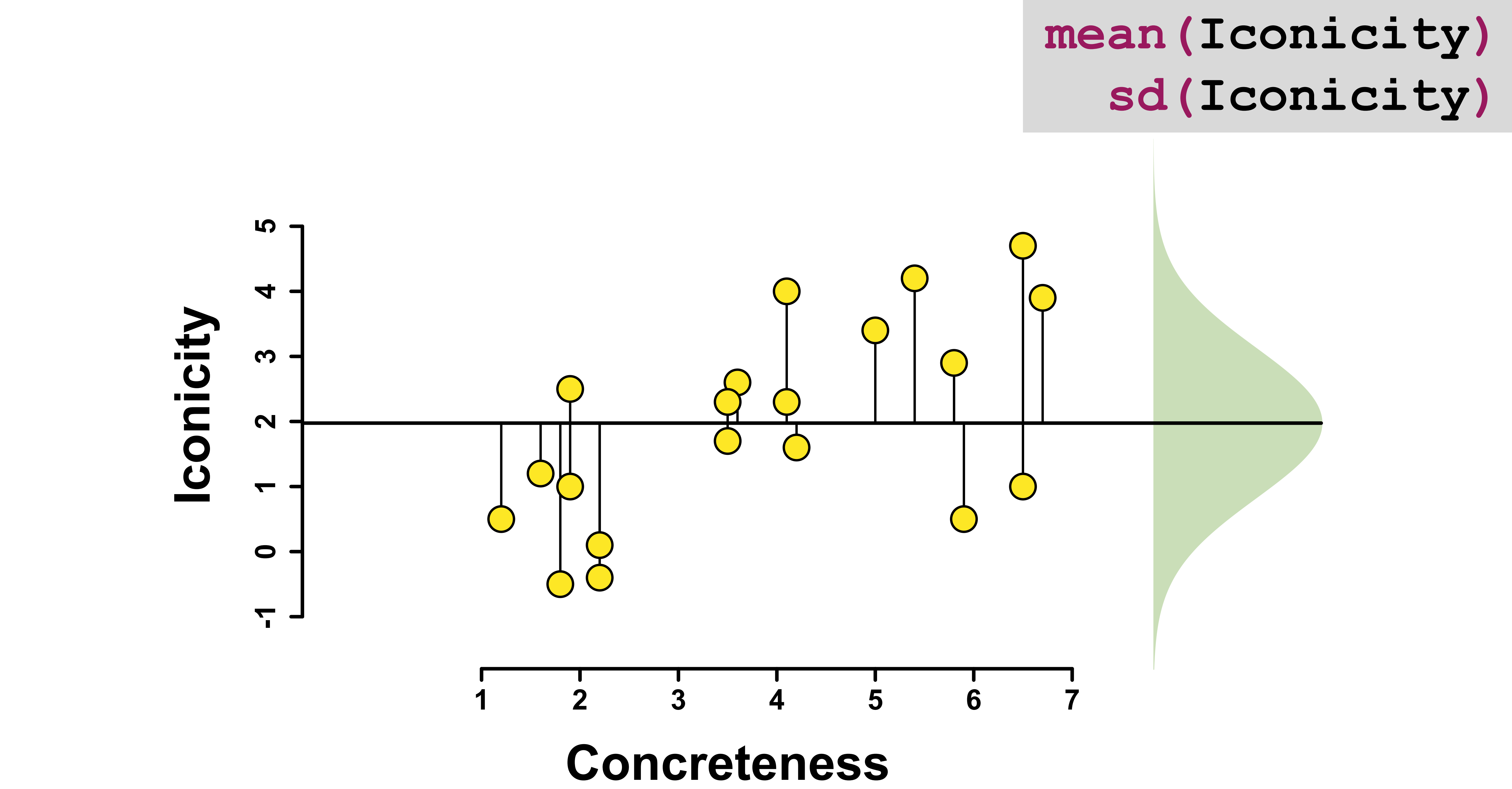

Linear models
Generalized Linear Models
Linear mixed effects models
Structural Equation Models
Generalized Additive Models


Winter, Perlman, Perry, & Lupyan (2017), *Interaction Studies*


Iconicity
-1
0
1
2
3
4
5
1
2
3
4
5
6
7
Concreteness

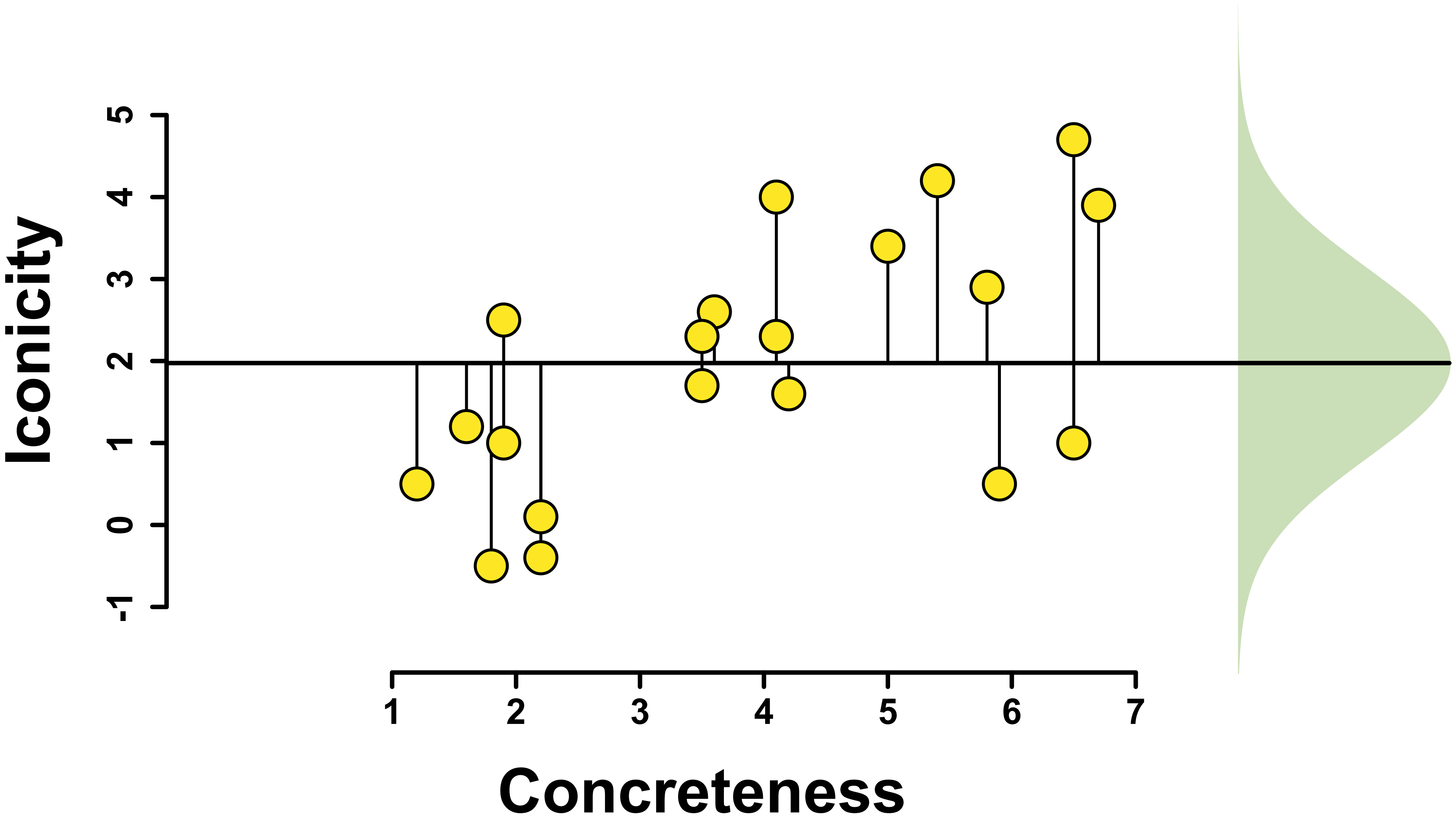
```
mean(Iconicity)
sd(Iconicity)
```





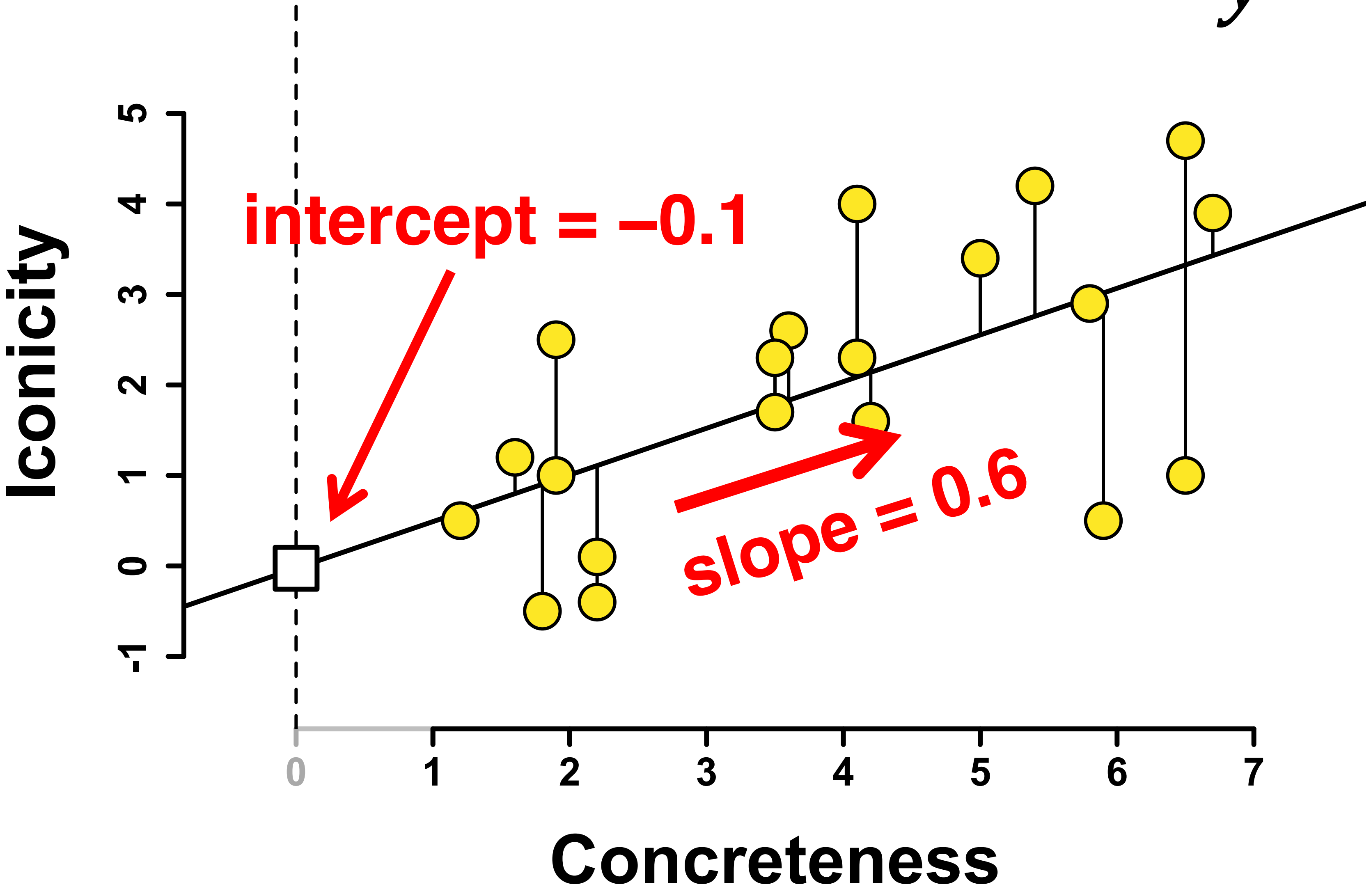
mean(Iconicity)
sd(Iconicity)
Iconicity
-1
0
1
2
3
4
5
1
2
3
4
5
6
7
Concreteness

```
lm(Iconicity ~ 1)
```



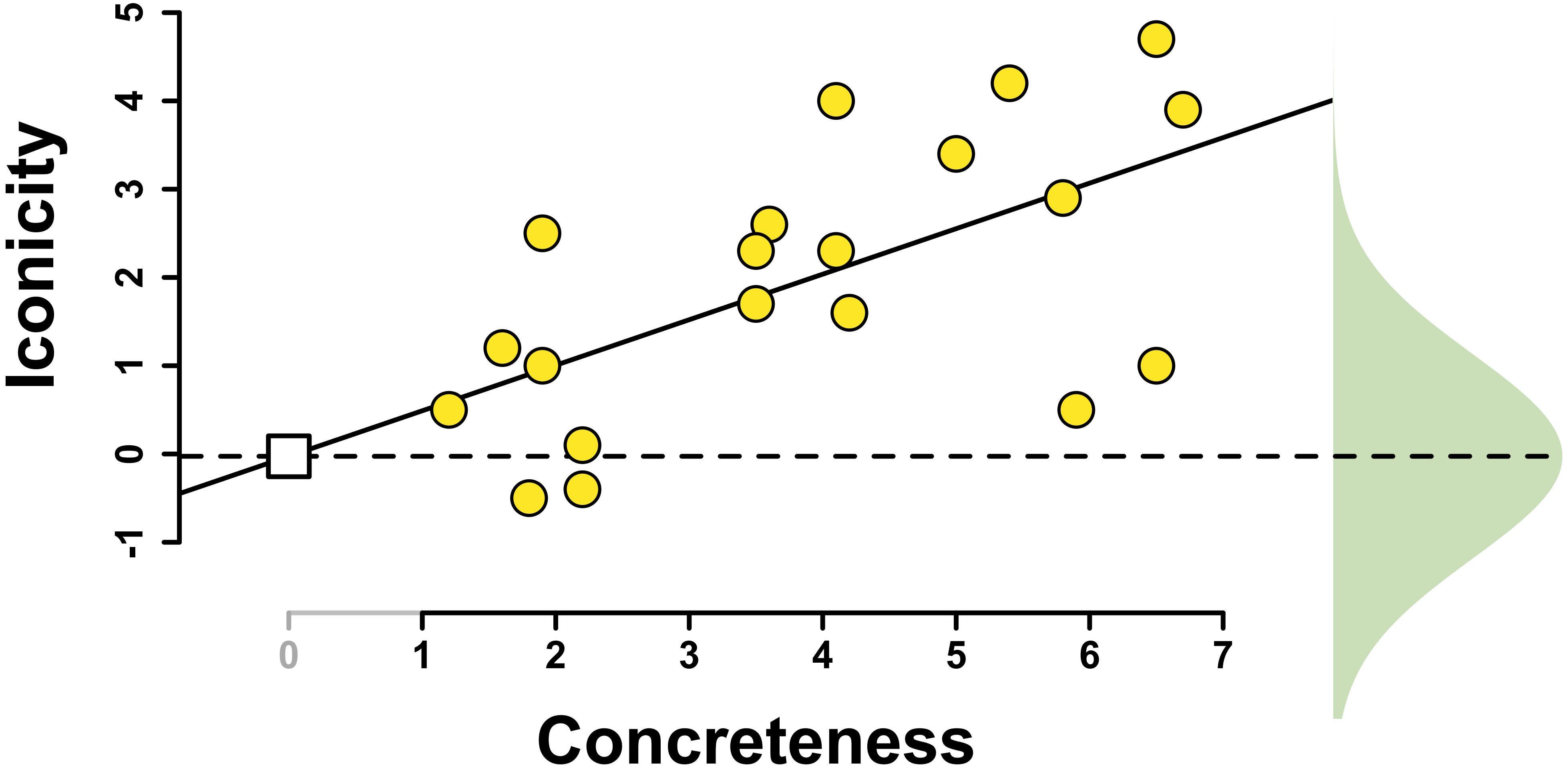
```
lm(Iconicity ~ 1 + Concreteness)
```

$$y = b_0 + b_1 * x$$



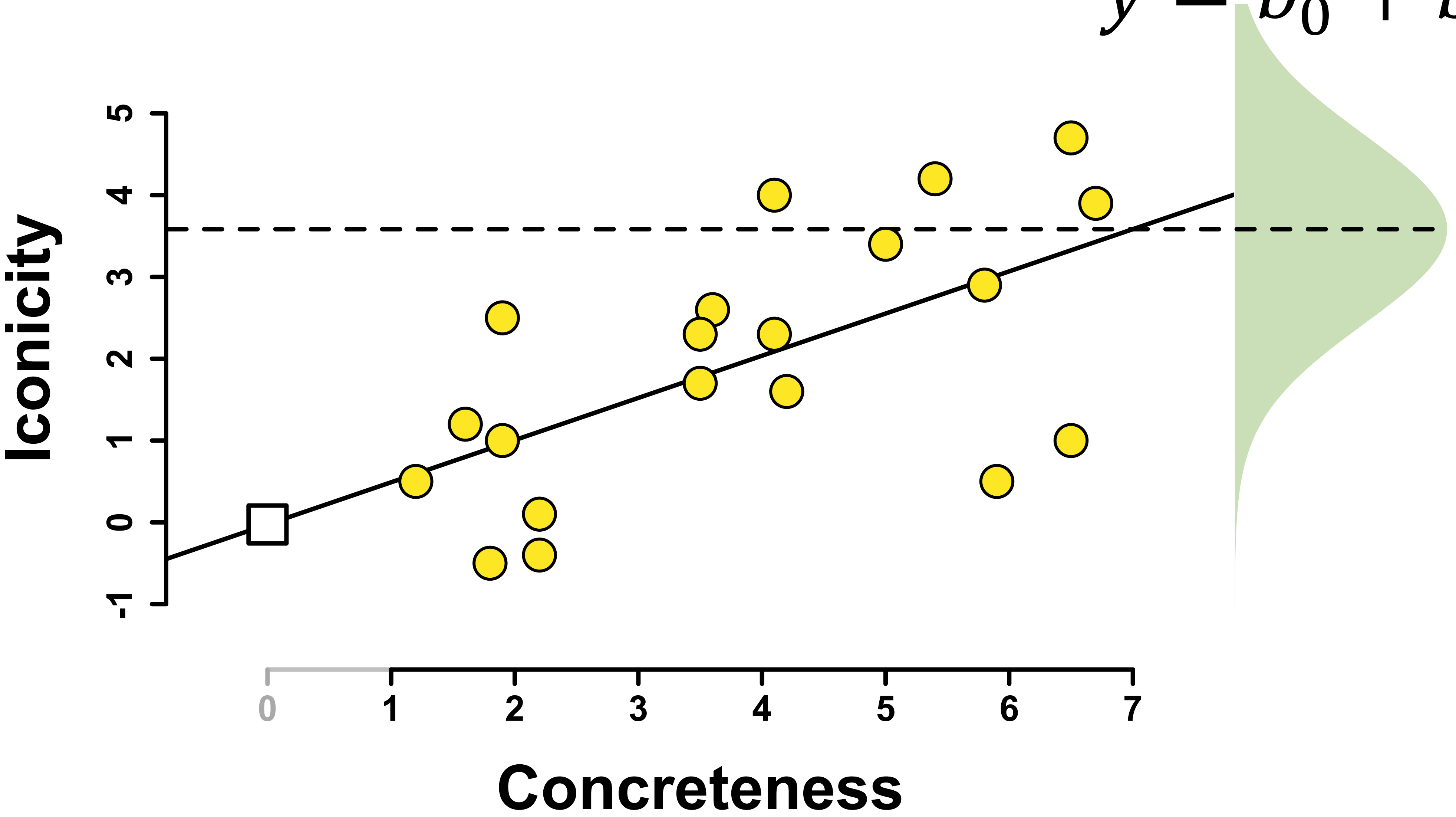
```
lm(Iconicity ~ Concreteness)
```

$$y = b_0 + b_1 * 0$$



`lm(Iconicity ~ Concreteness)`

$$y = b_0 + b_1 * 7$$



```
lm(y ~ x)
```

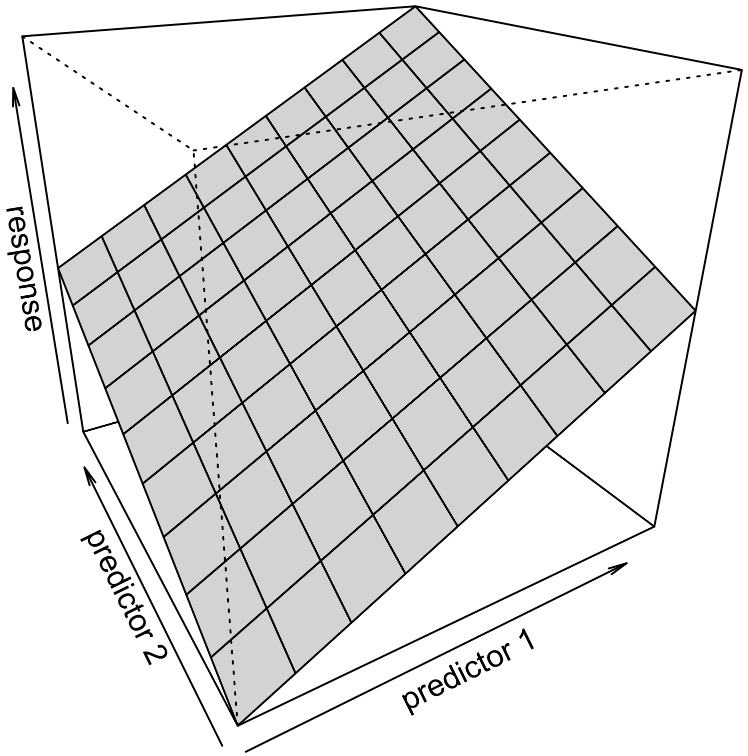
link function

linear predictor

$$I(\underbrace{\beta_0 + \beta_1 * x_i}_{\text{linear predictor}})$$

$$y_i \sim Normal(\mu_i, \sigma)$$

```
lm(y ~ x1 + x2)
```

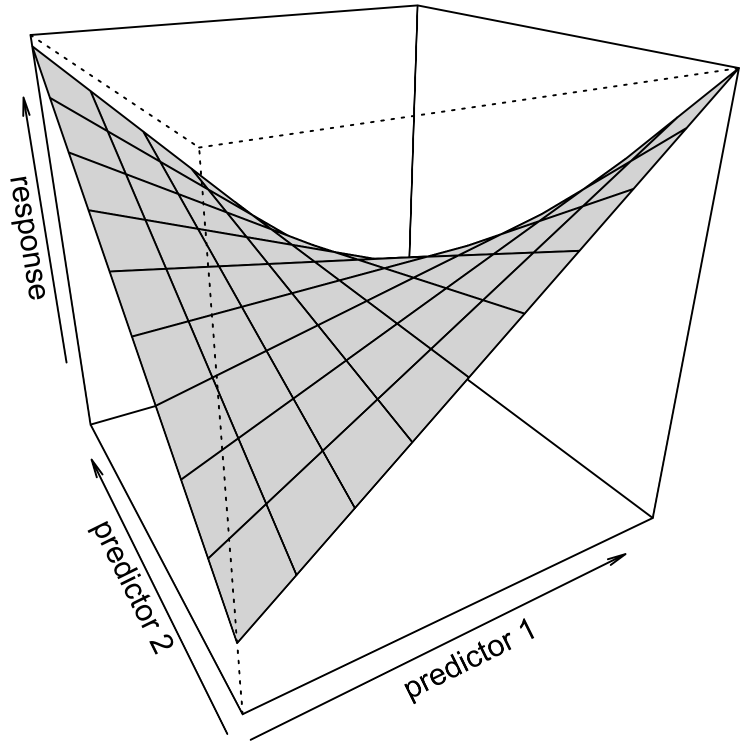



$$\beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$y \sim Normal(\mu, \sigma)$$

```
lm(y ~ x1 + x2 + x1:x2)
```

```
lm(y ~ x1 * x2)
```




$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \overbrace{\beta_3 (x_1 x_2)}^{\textbf{interaction term}}$$

$$y \sim Normal(\mu, \sigma)$$

# A bit more on the concept of an interaction

```
lm(plant growth ~ sun exposure * water)
```

~

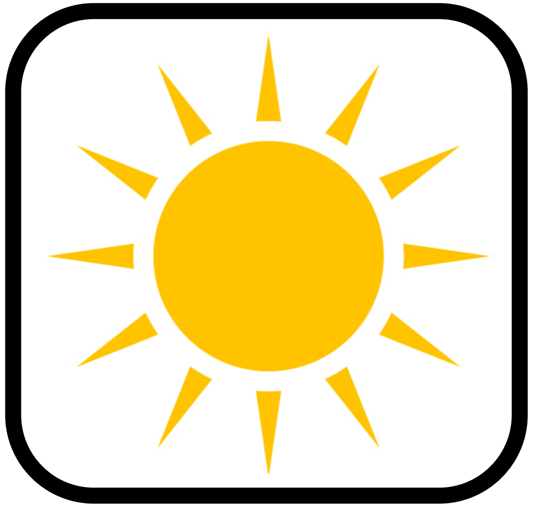
*

example from McElreath (2016), *Statistical Rethinking*

# A bit more on the concept of an interaction

```
lm(plant growth ~ sun exposure * water)
```

~



*

example from McElreath (2016), *Statistical Rethinking*

# A bit more on the concept of an interaction

```
lm(plant growth ~ sun exposure * water)
```

~

*

example from McElreath (2016), *Statistical Rethinking*

**Exercise with `broom`**

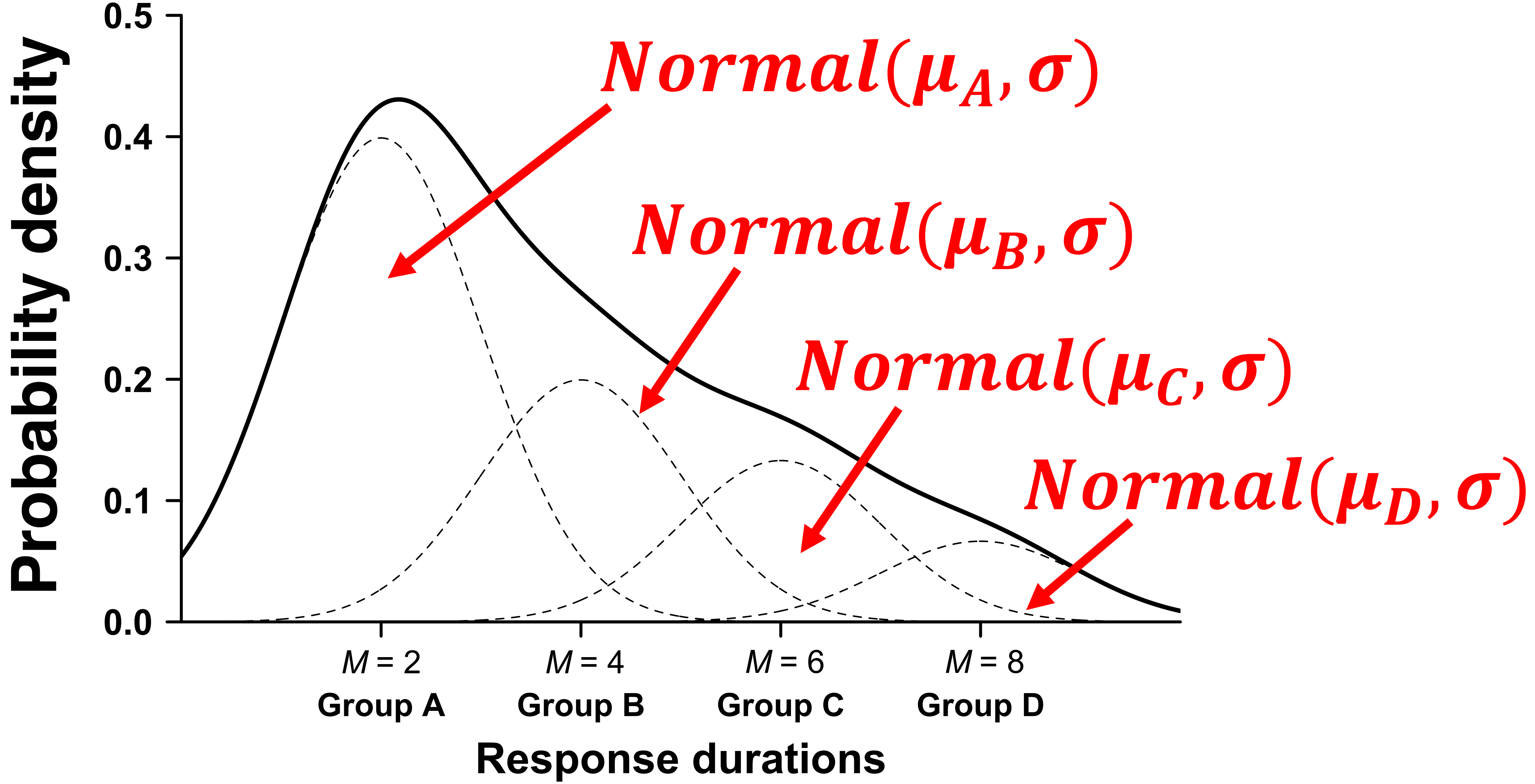
Linear models 101

Intro to GLM

Exercise with `broom`

Linear models 101

Intro to GLM



Probability density
0.5
0.4
0.3
0.2
0.1
0.0
$Normal(\mu_A, \sigma)$
$Normal(\mu_B, \sigma)$
$Normal(\mu_C, \sigma)$
$Normal(\mu_D, \sigma)$
*M* = 2
Group A
*M* = 4
Group B
*M* = 6
Group C
*M* = 8
Group D
Response durations

```
lm(y ~ x)
```

link function

$$I(\overbrace{\beta_0 + \beta_1 * x_i}^{\text{linear predictor}})$$

$$y_i \sim Normal(\mu_i, \sigma)$$

`glm(y ~ x, family = ...)`

`family = poisson`

**Poisson distribution for count data**

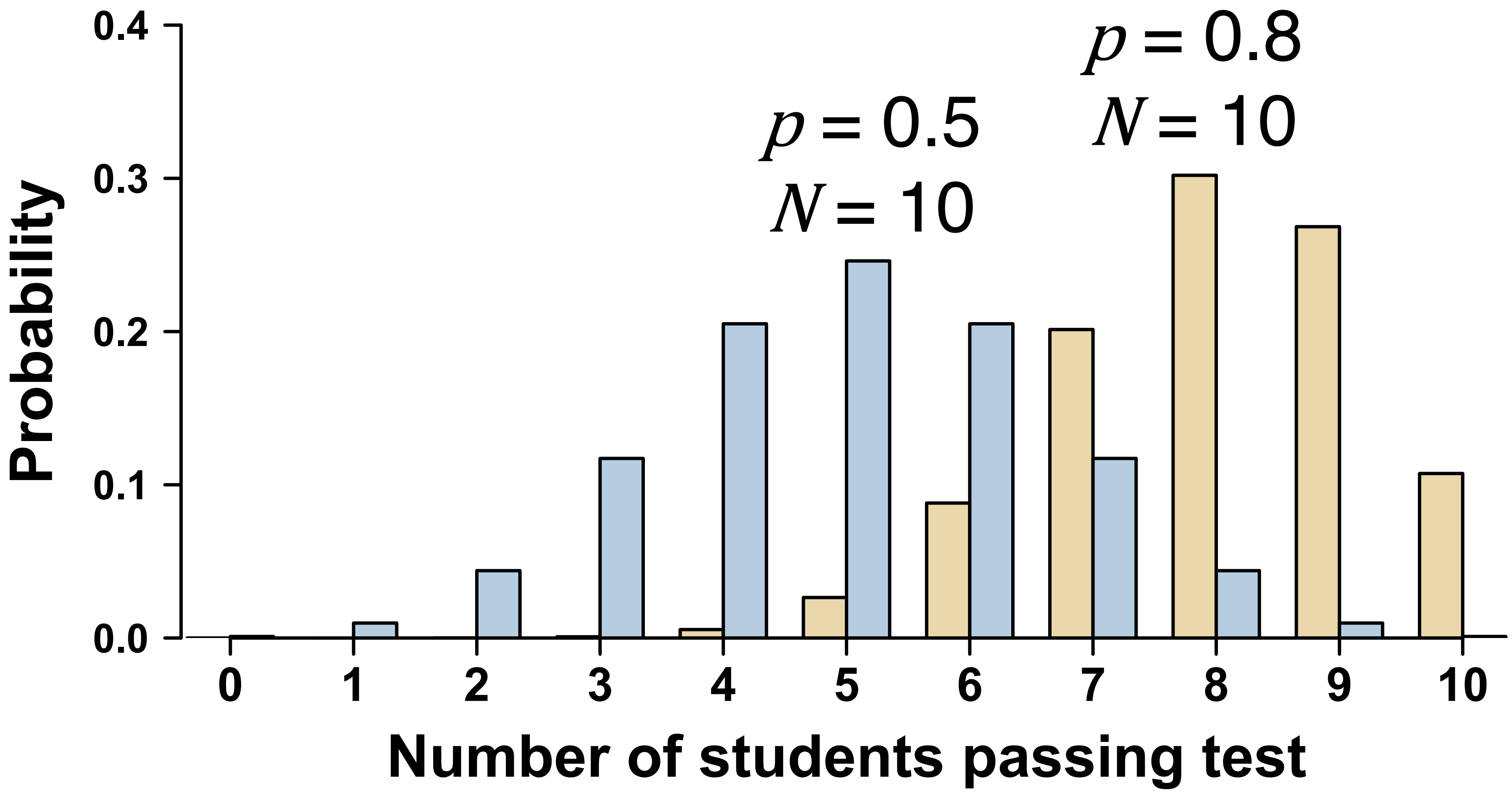
`family = binomial`

**Bernoulli distribution for binary data**

$$y_i \sim Zaphod(\theta_i)$$

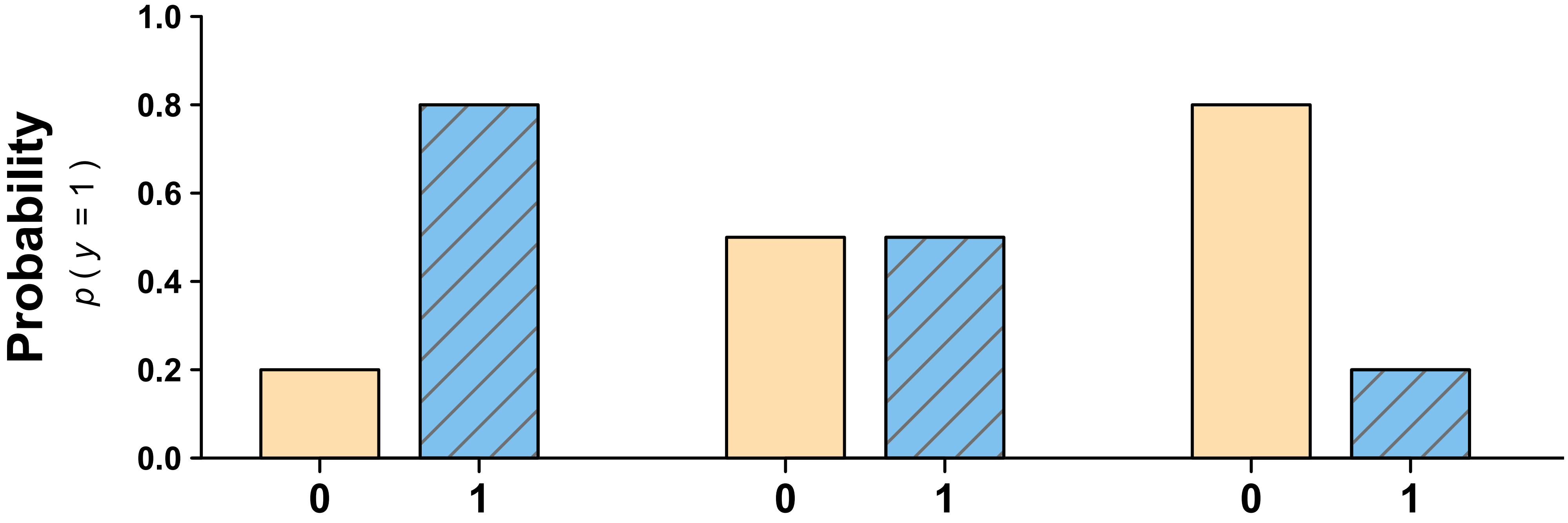
# The binomial distribution

**two parameters:**
***p*, *N***



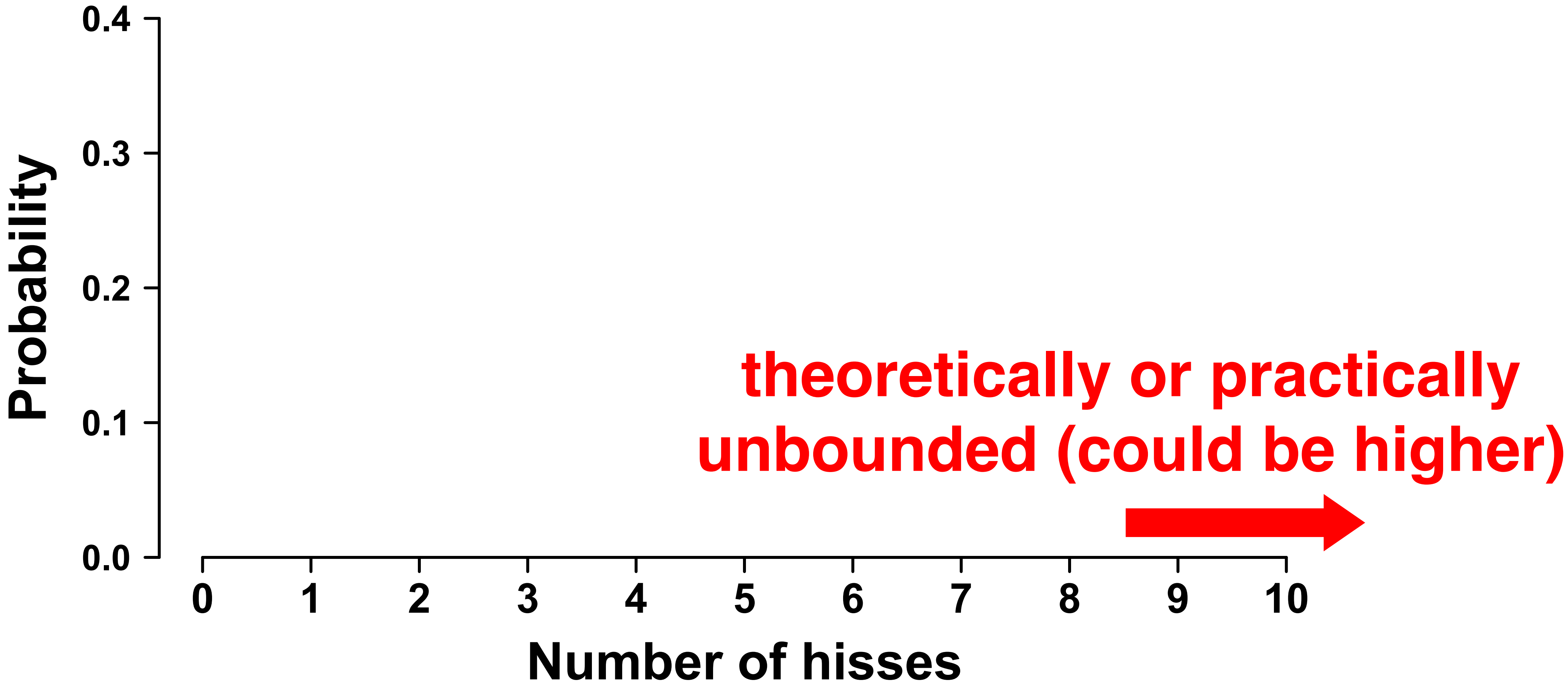
fix $N$ to 1

# The Bernoulli distribution

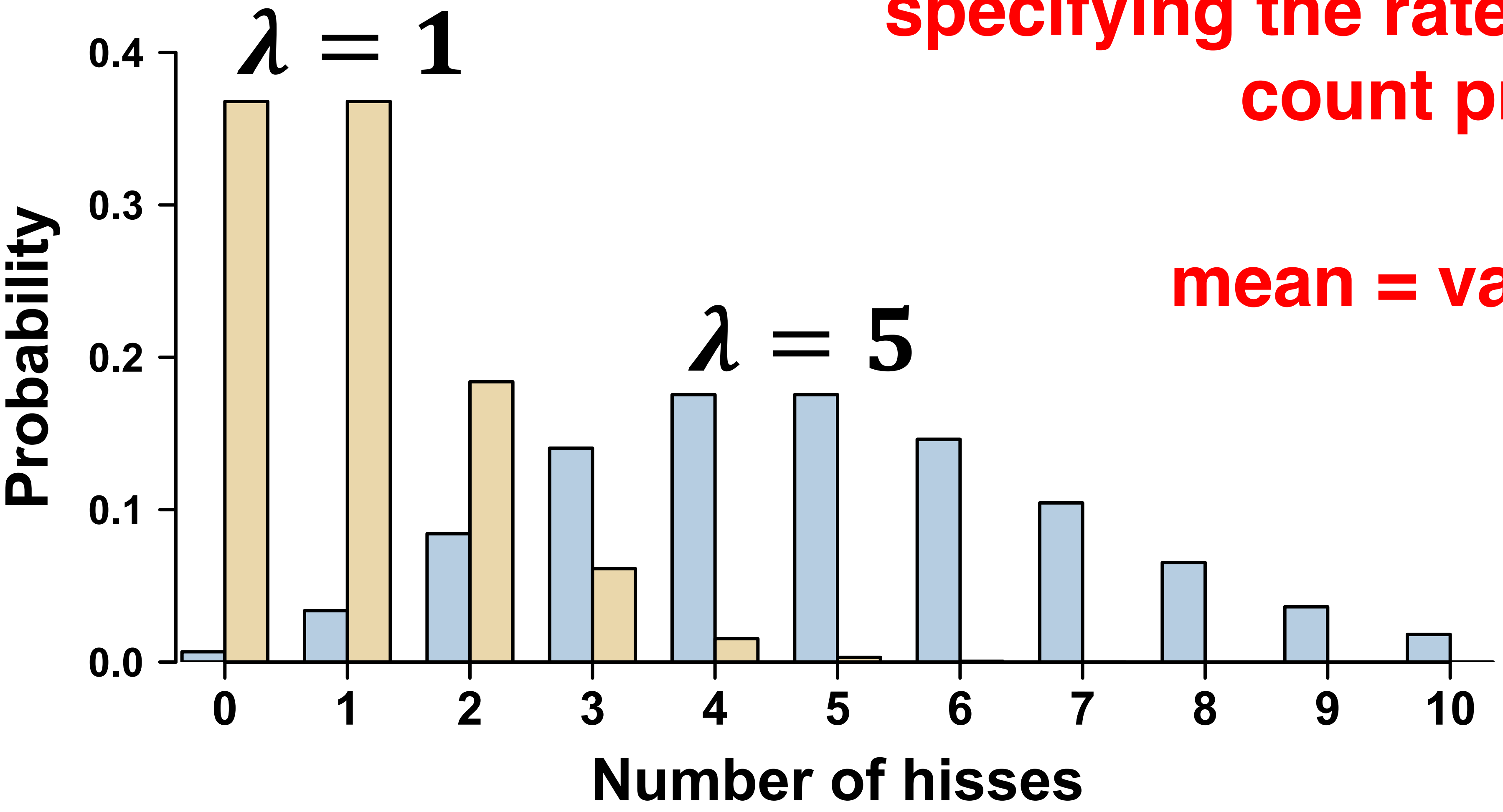



one parameter: $p$, specifying the probability of occurrence

# The Poisson distribution



# The Poisson distribution




one parameter: λ, specifying the rate of the count process

mean = variance

# Poisson regression

$$\beta_0 + \beta_1 * x_i$$

$$\downarrow$$

$$y_i \sim Poisson(\lambda_i)$$

# Poisson regression

$$e^{\beta_0+\beta_1 * x_i}$$

$$y_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 * x_i$$

coefficients are
in log space

# Logistic regression

$$logit^{-1}(\beta_0 + \beta_1 * x_i)$$

$$\downarrow$$

$$y_i \sim Bernoulli(p_i)$$

$$\mathrm{logit}(p_i) = \beta_0 + \beta_1 * x_i$$

coefficients are
in log odd space

# Overview: some standard generalized linear models

$I(\beta_0 + \beta_1 * x_i)$ $\downarrow$ $y_i \sim Normal(\mu_i, \sigma)$

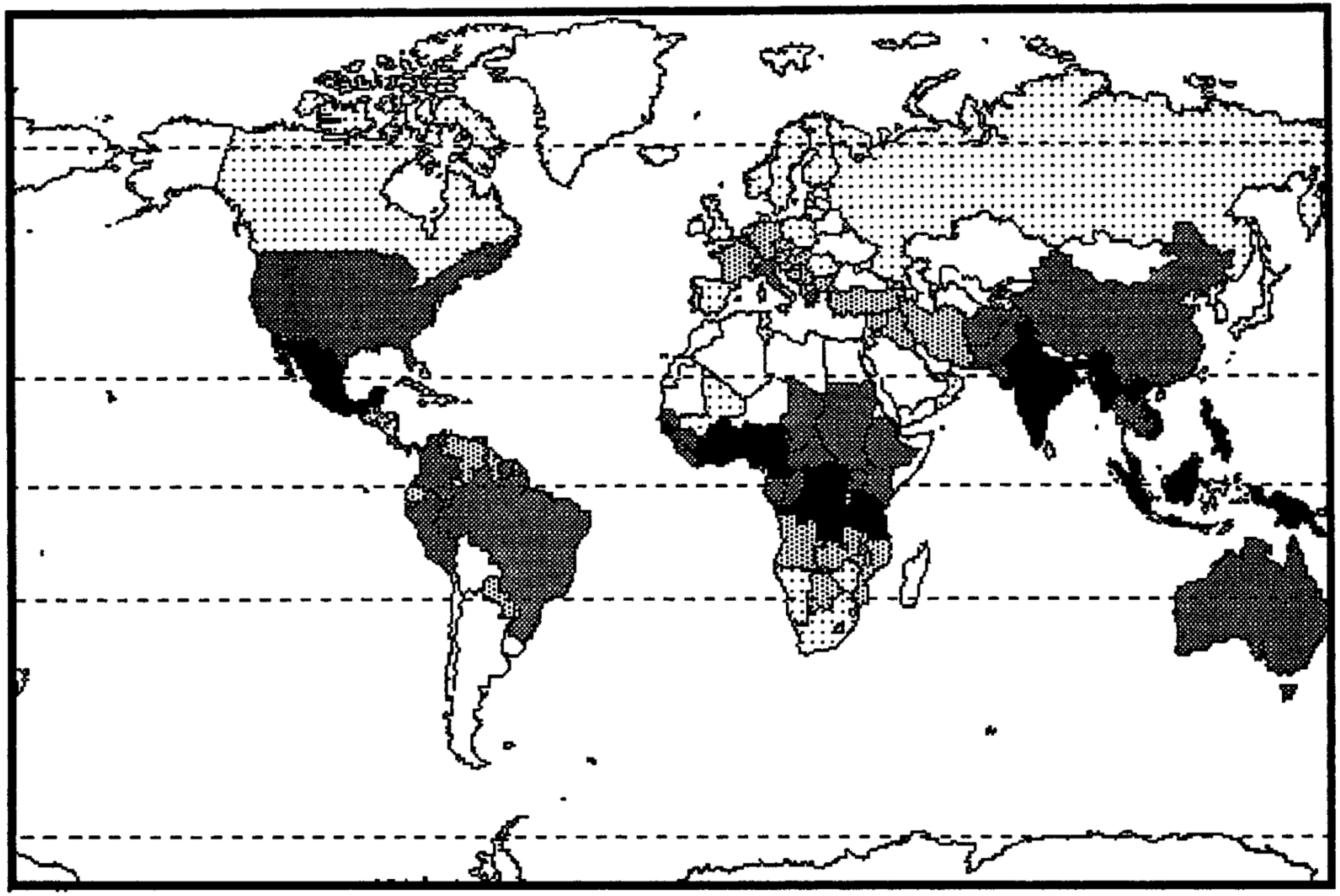
**Linear regression**

$logistic(\beta_0 + \beta_1 * x_i)$ $\downarrow$ $y_i \sim Bernoulli(p_i)$

**Logistic regression**

$exp(\beta_0 + \beta_1 * x_i)$ $\downarrow$ $y_i \sim Poisson(\lambda_i)$

**Poisson regression**



Nettle, D. (1999). *Linguistic diversity*. Oxford, UK: Oxford University Press.

# Intro to linear models and generalized linear models

Bodo Winter

UNIVERSITY OF BIRMINGHAM

@BodoWinter

www.bodowinter.com