Methods in Psycholinguistics

— Mixed effects logistic regression —

Judith Degen
Stanford University

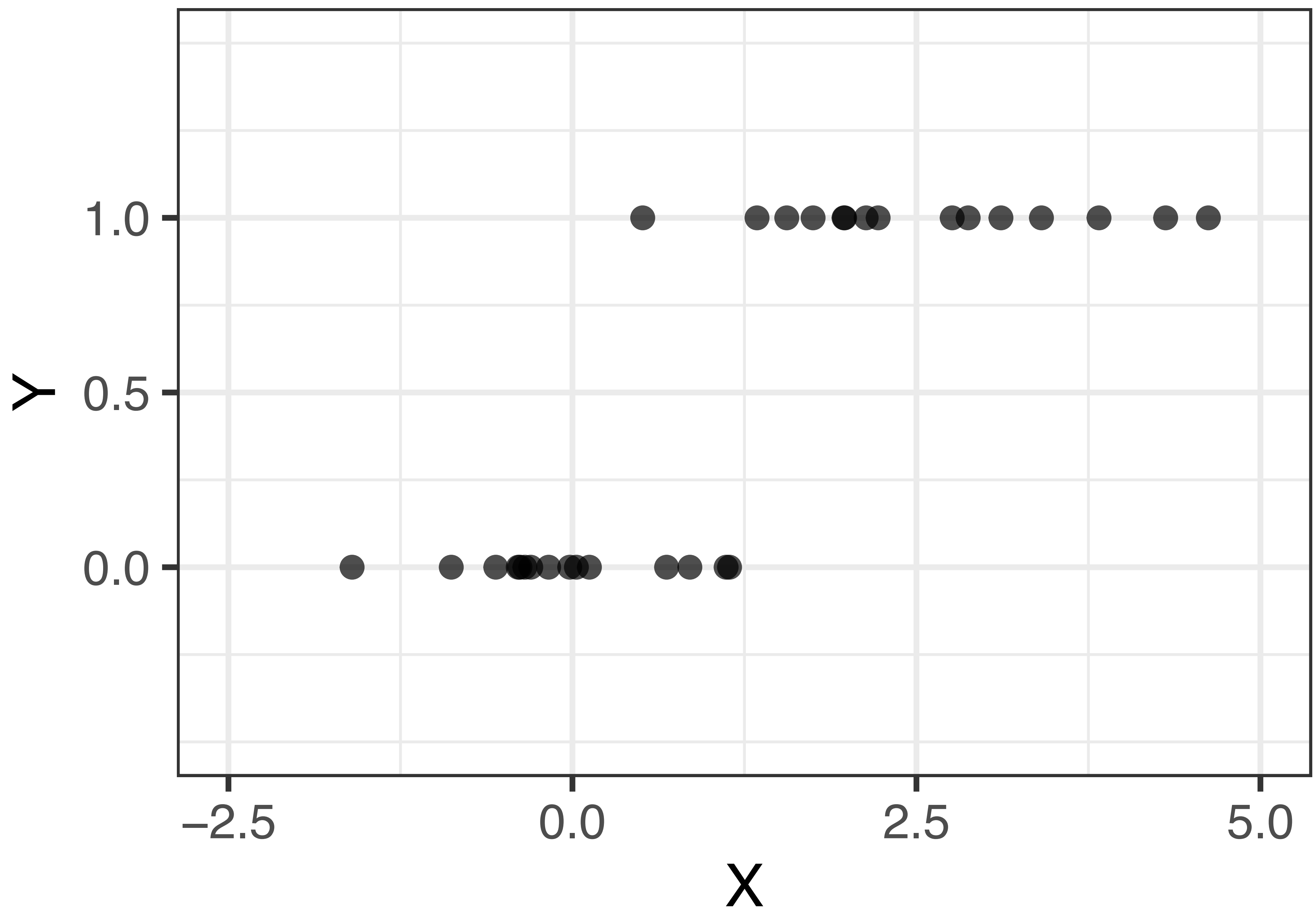
Logistic regression

- for binary (categorical) instead of continuous outcomes
- instead of predicting the mean of an outcome, we’re predicting the log odds of an event occurring
- also called “logit model”

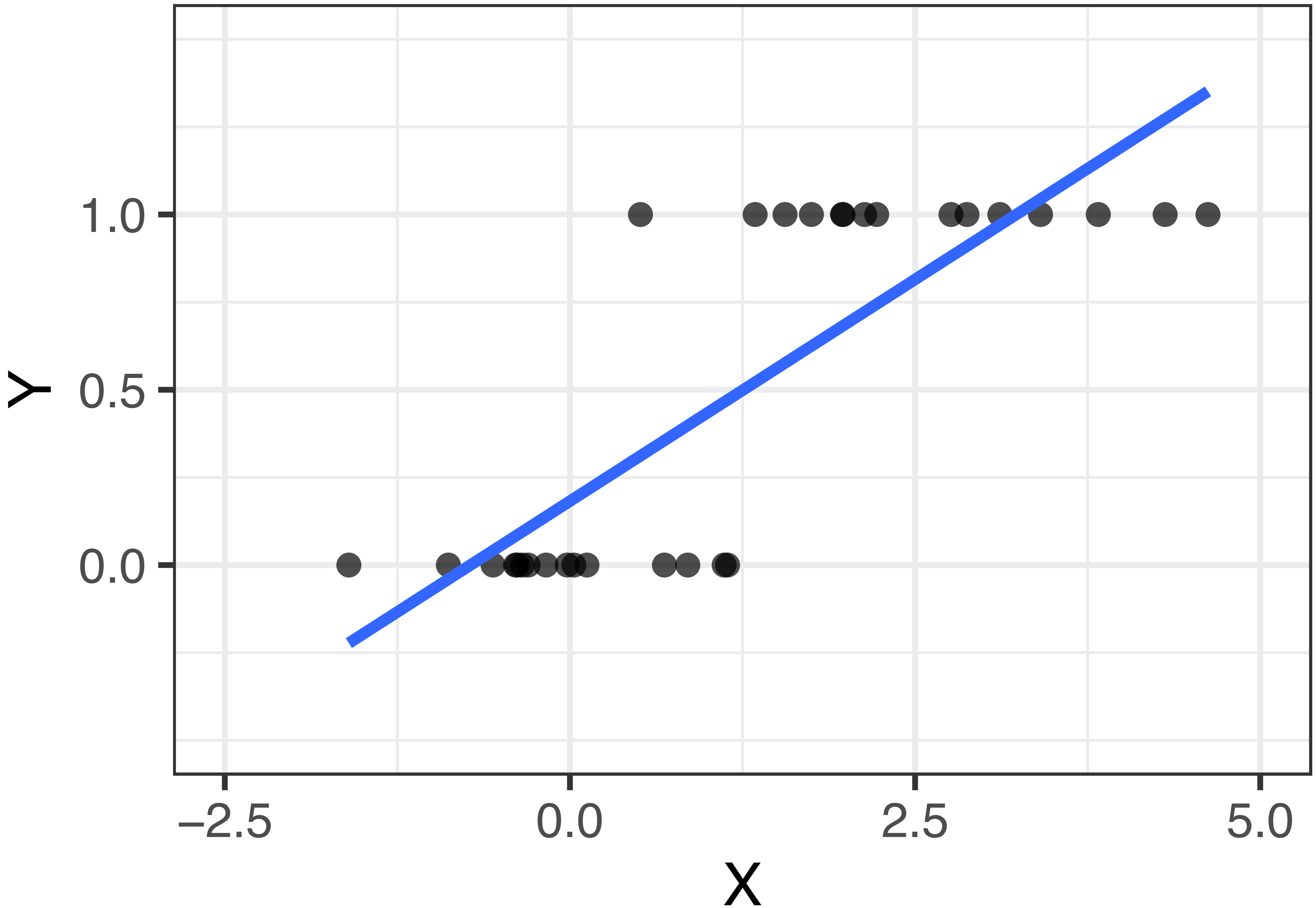
What kind of data?

- grammaticality (binary)
- syntactic variation (e.g., dative alternation)
- phonological variation (e.g., t-deletion)
- experimental forced choice (eg., truth-value judgments)
- eye-tracking data (e.g., look to target)

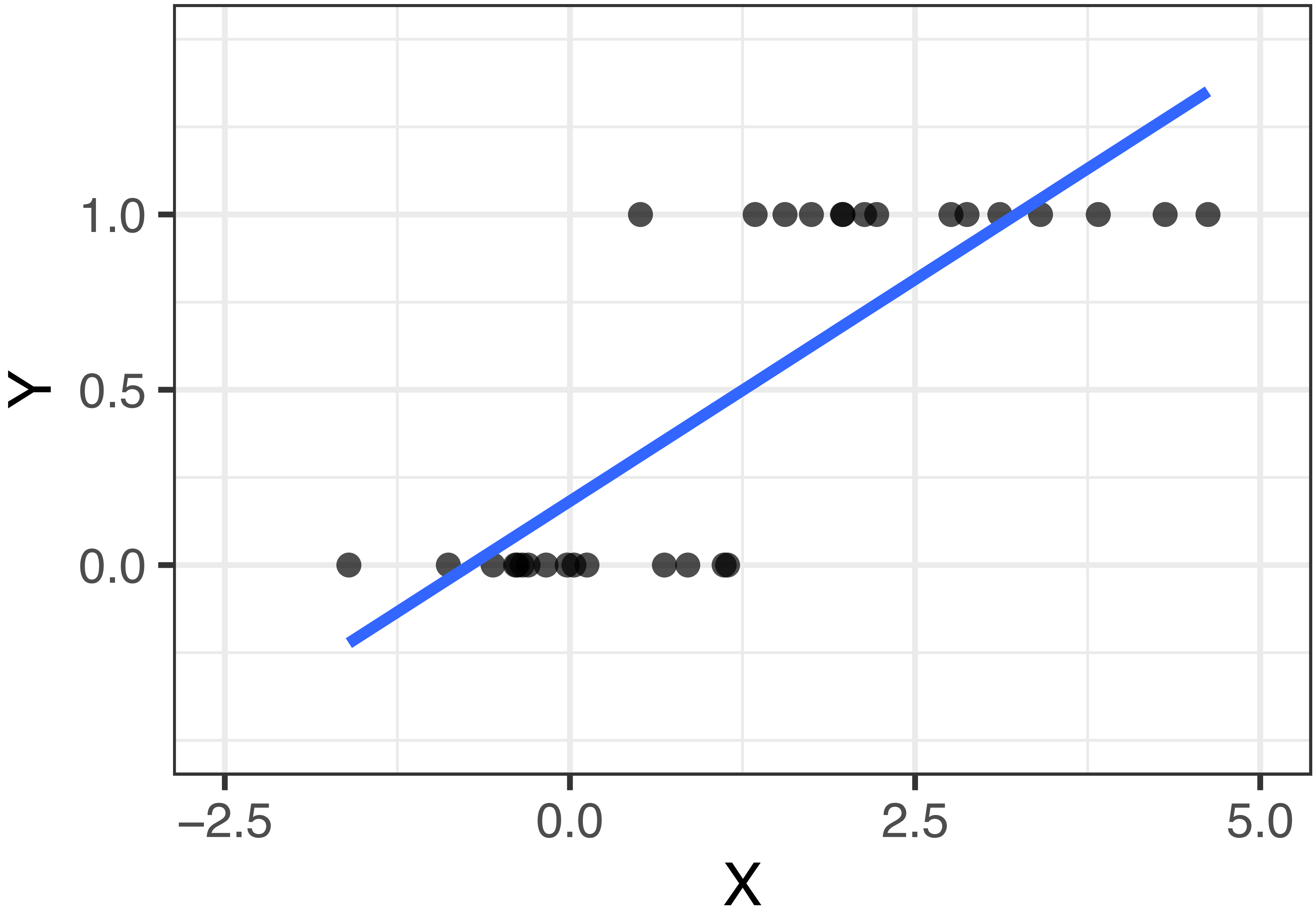
Why not use linear regression for categorical outcomes?



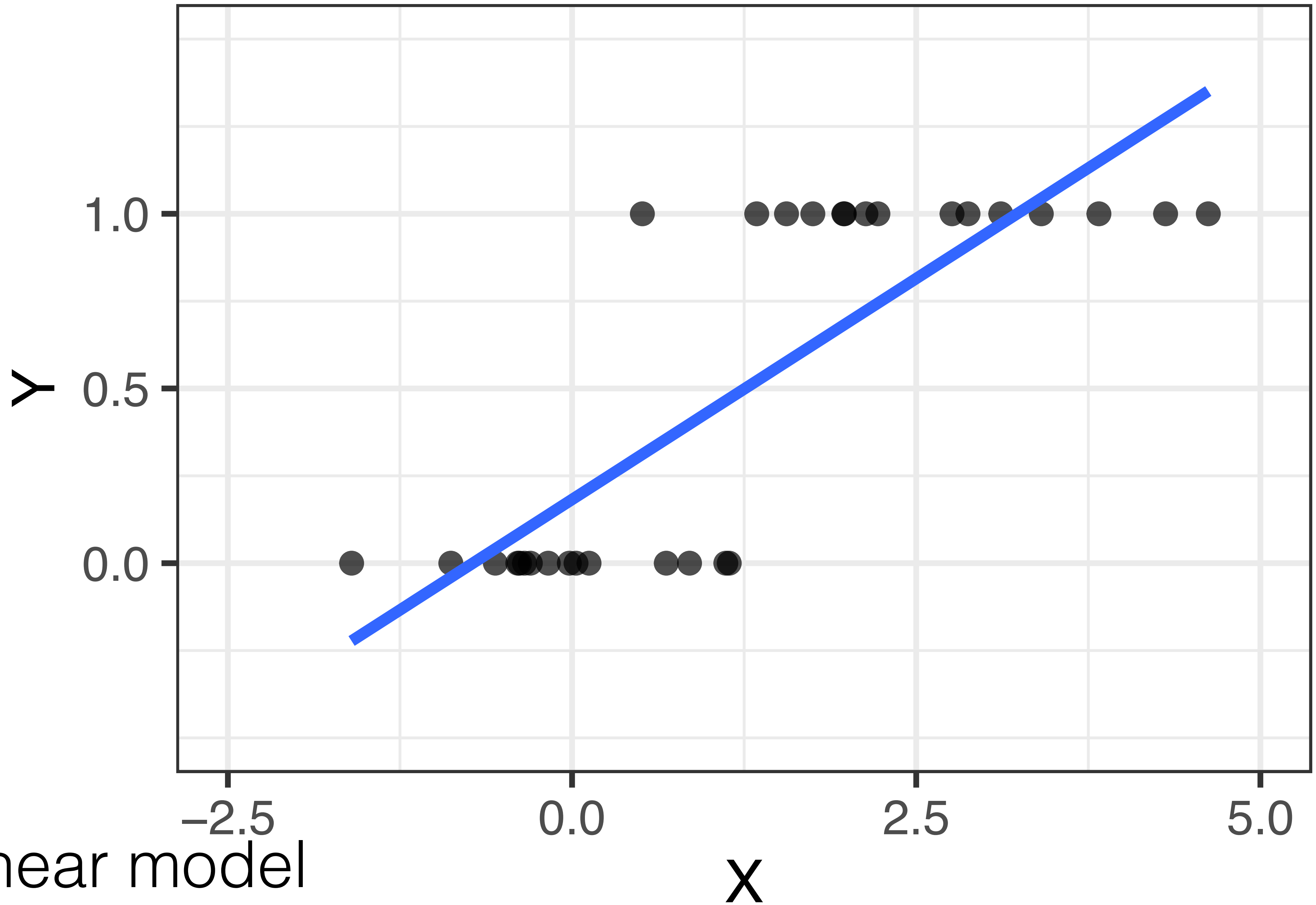
Why not use linear regression for categorical outcomes?



Why not use linear regression for categorical outcomes?



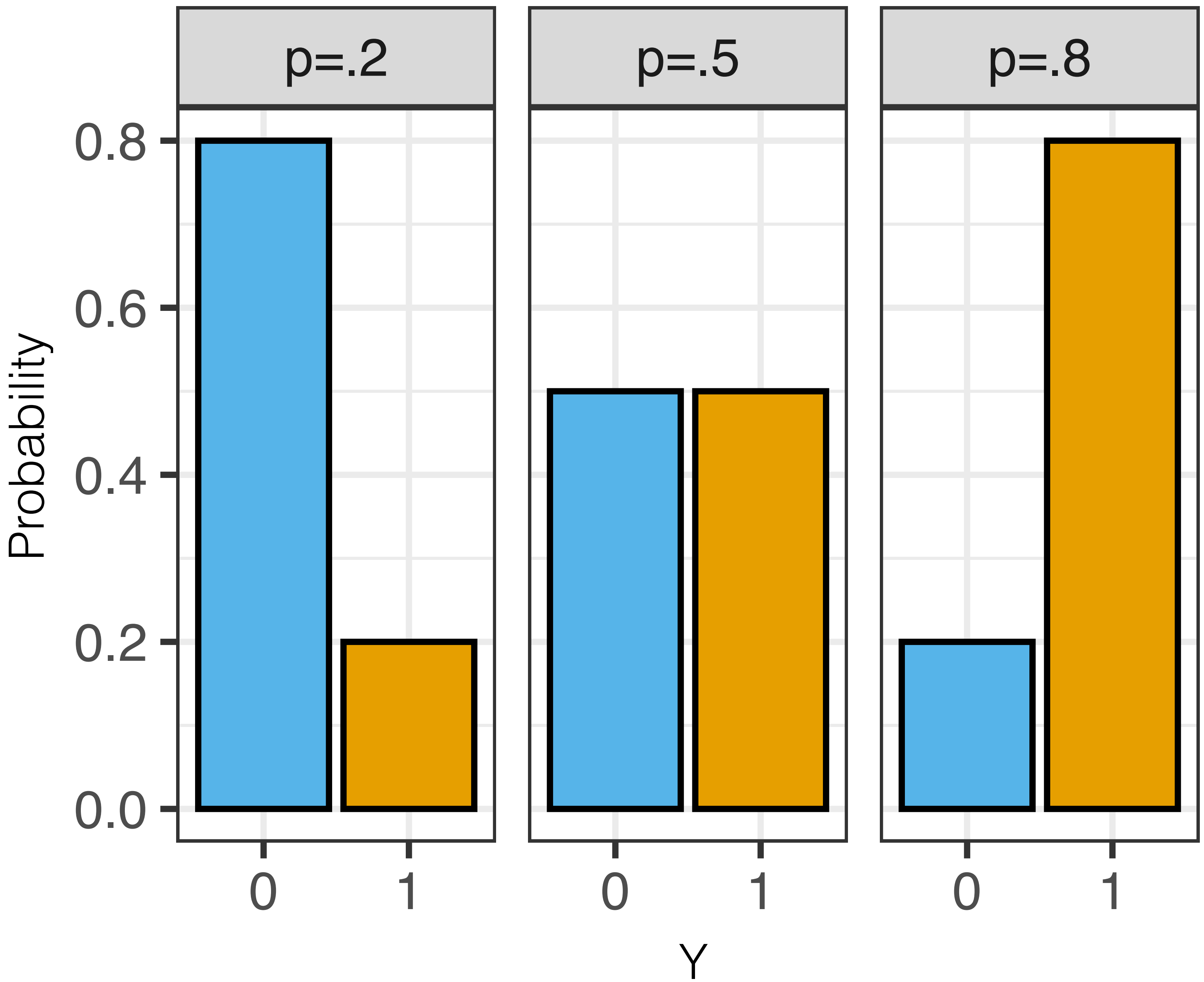
Why not use linear regression for categorical outcomes?



The linear model

- makes impossible predictions (values of Y > 1 or Y < 0)
- is meaningless if its assumptions are violated

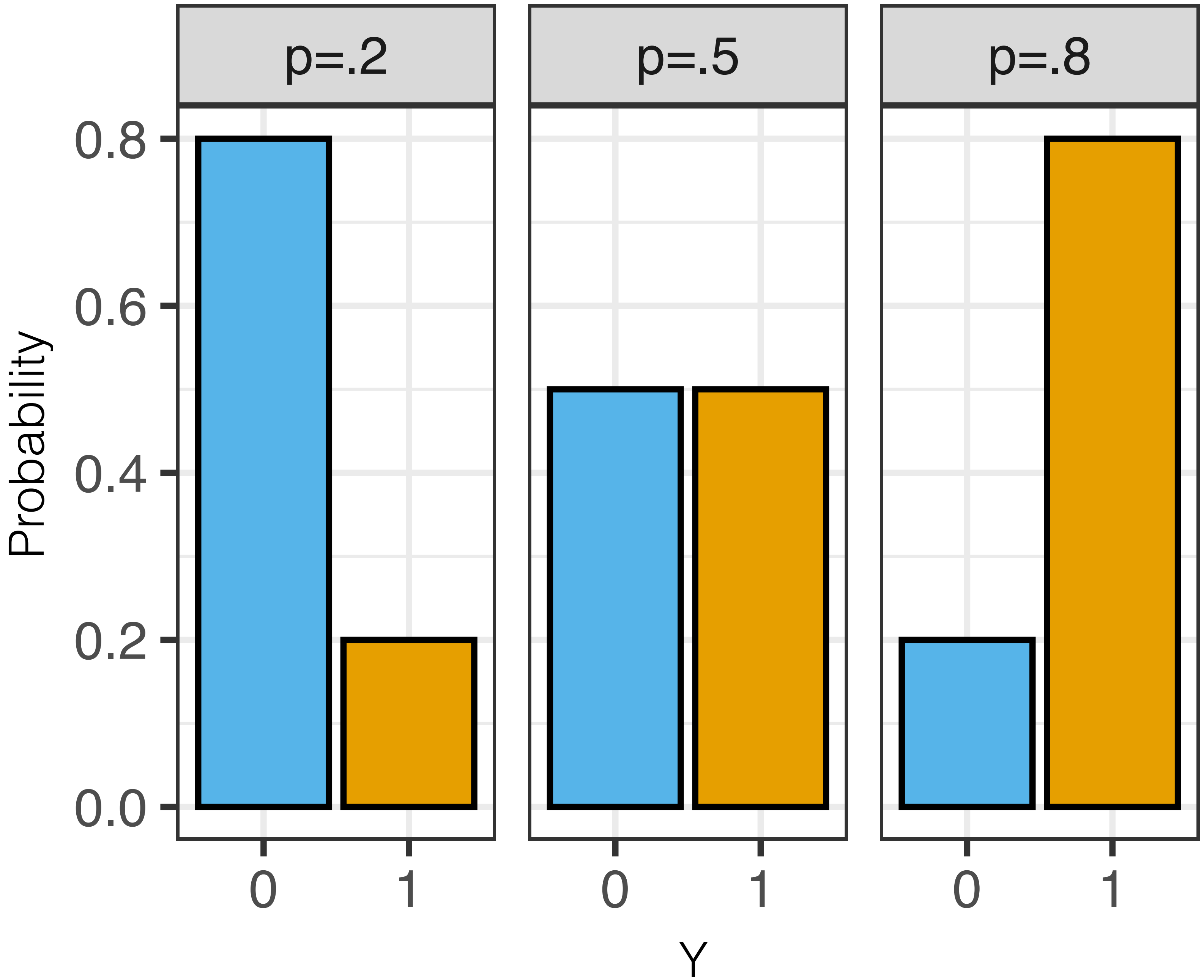
Logistic regression



Assume outcome is Bernoulli-distributed (special case of binomial distribution)

$$y_i \sim \text{Bernoulli}(p_i)$$

Logistic regression



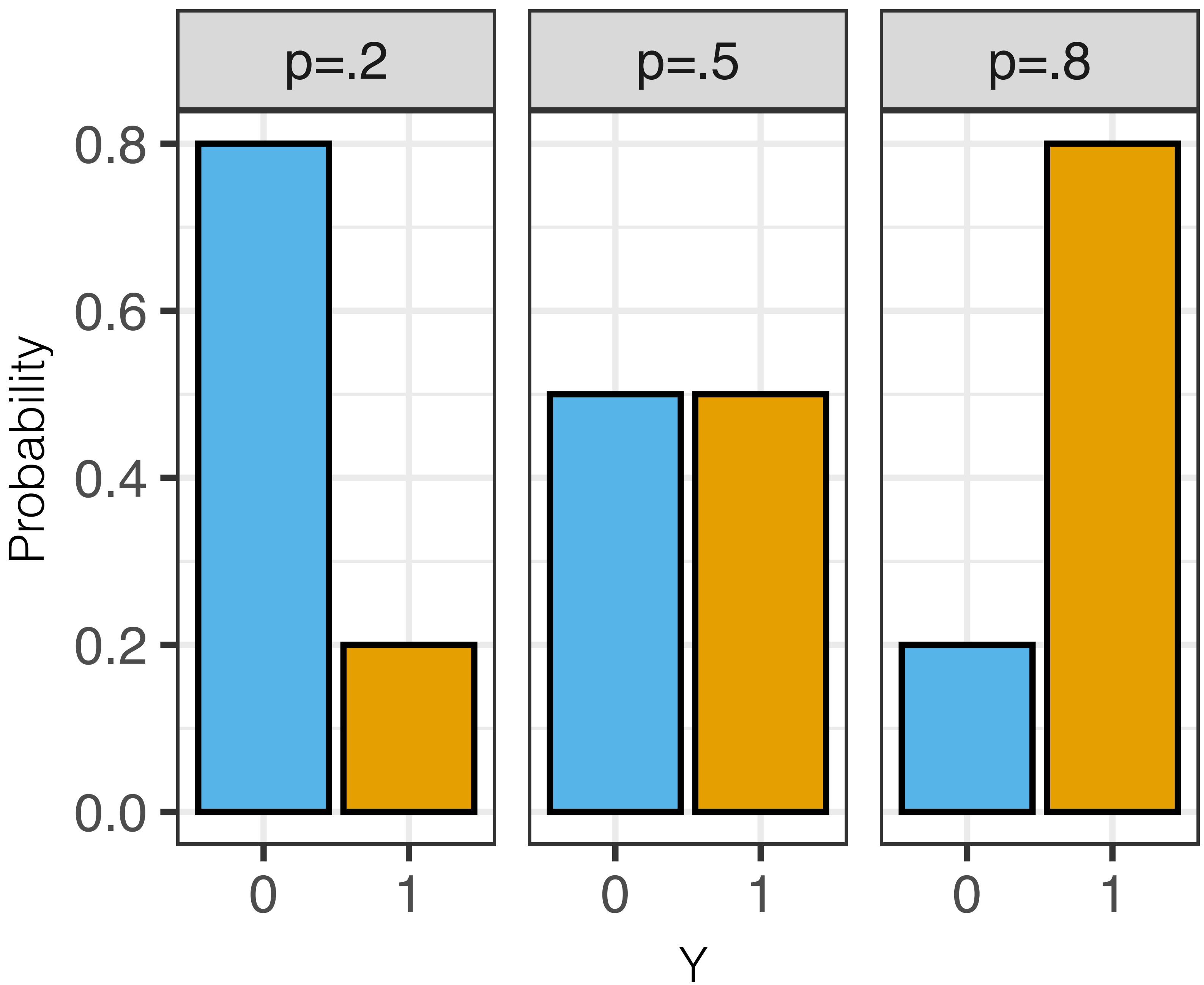
Assume outcome is Bernoulli-distributed (special case of binomial distribution)

$$y_i \sim \text{Bernoulli}(p_i)$$

So, ideally:

$$p_i = \beta_0 + \beta_1 x_i$$

Logistic regression



Assume outcome is Bernoulli-distributed (special case of binomial distribution)

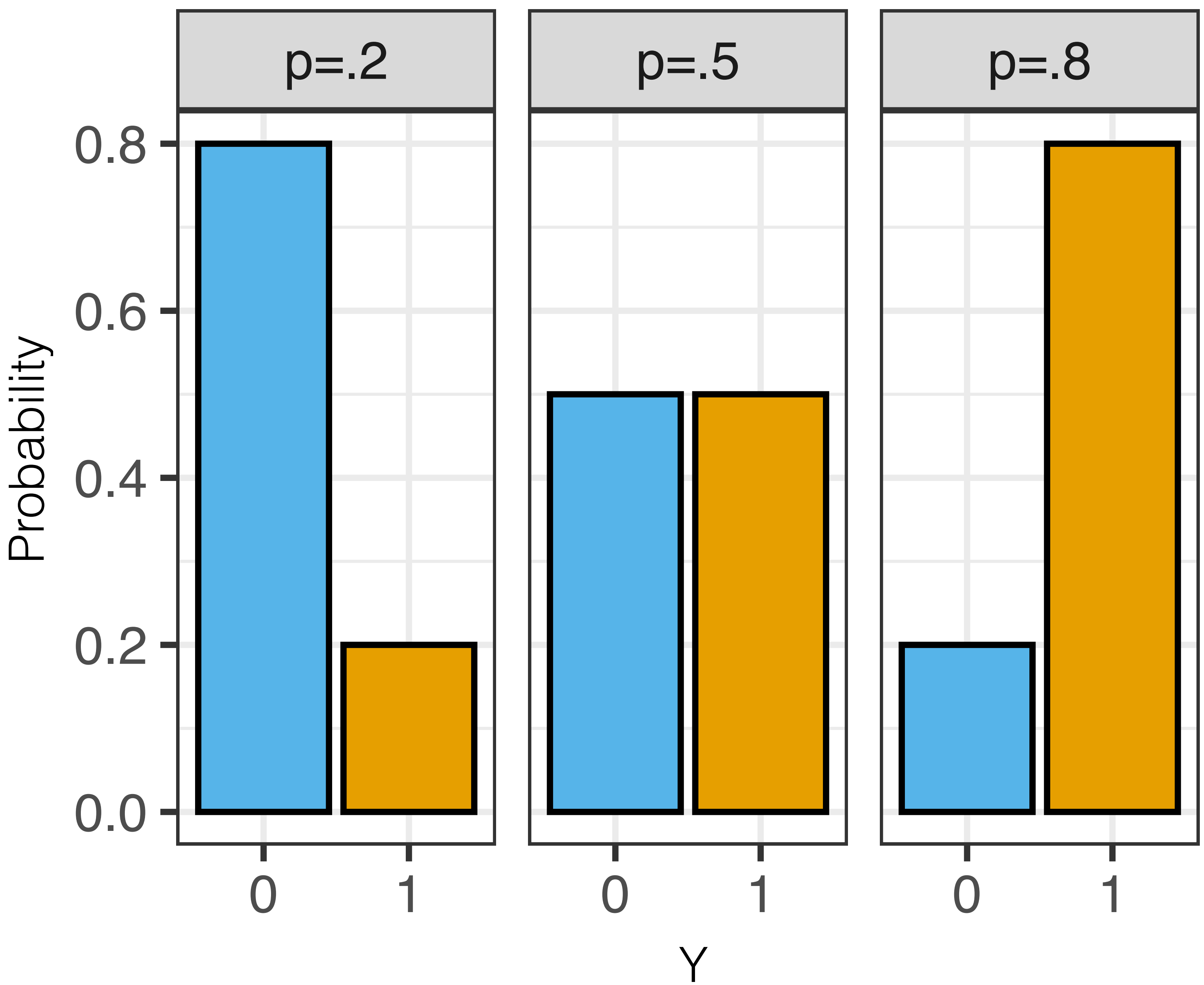
$$y_i \sim \text{Bernoulli}(p_i)$$

So, ideally:

$$p_i = \beta_0 + \beta_1 x_i$$

Why can't we do that?

Logistic regression



Assume outcome is Bernoulli-distributed (special case of binomial distribution)

$$y_i \sim \text{Bernoulli}(p_i)$$

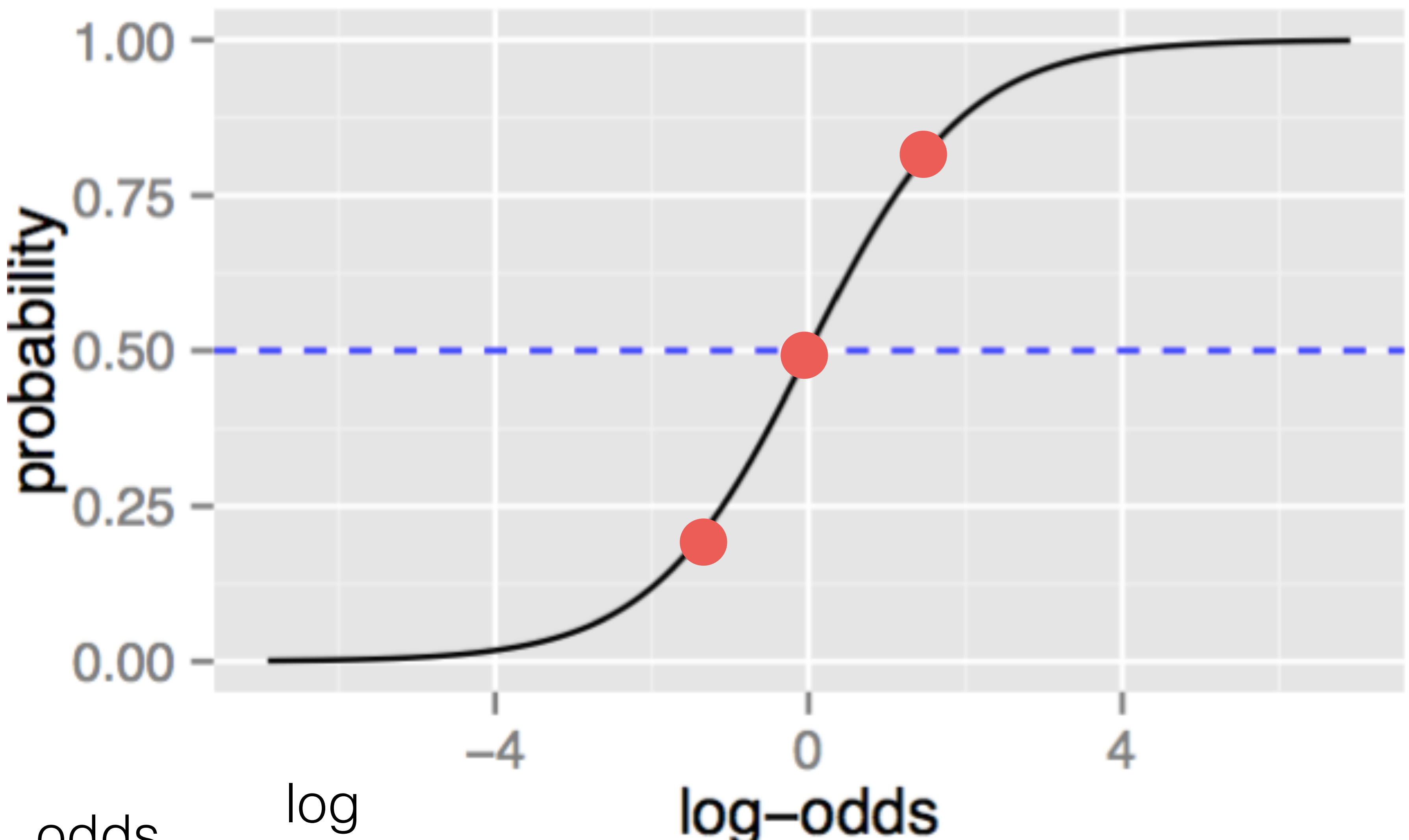
So, ideally:

$$p_i = \beta_0 + \beta_1 x_i$$

Why can't we do that?

Probabilities are bounded by [0,1]

Log odds and probability



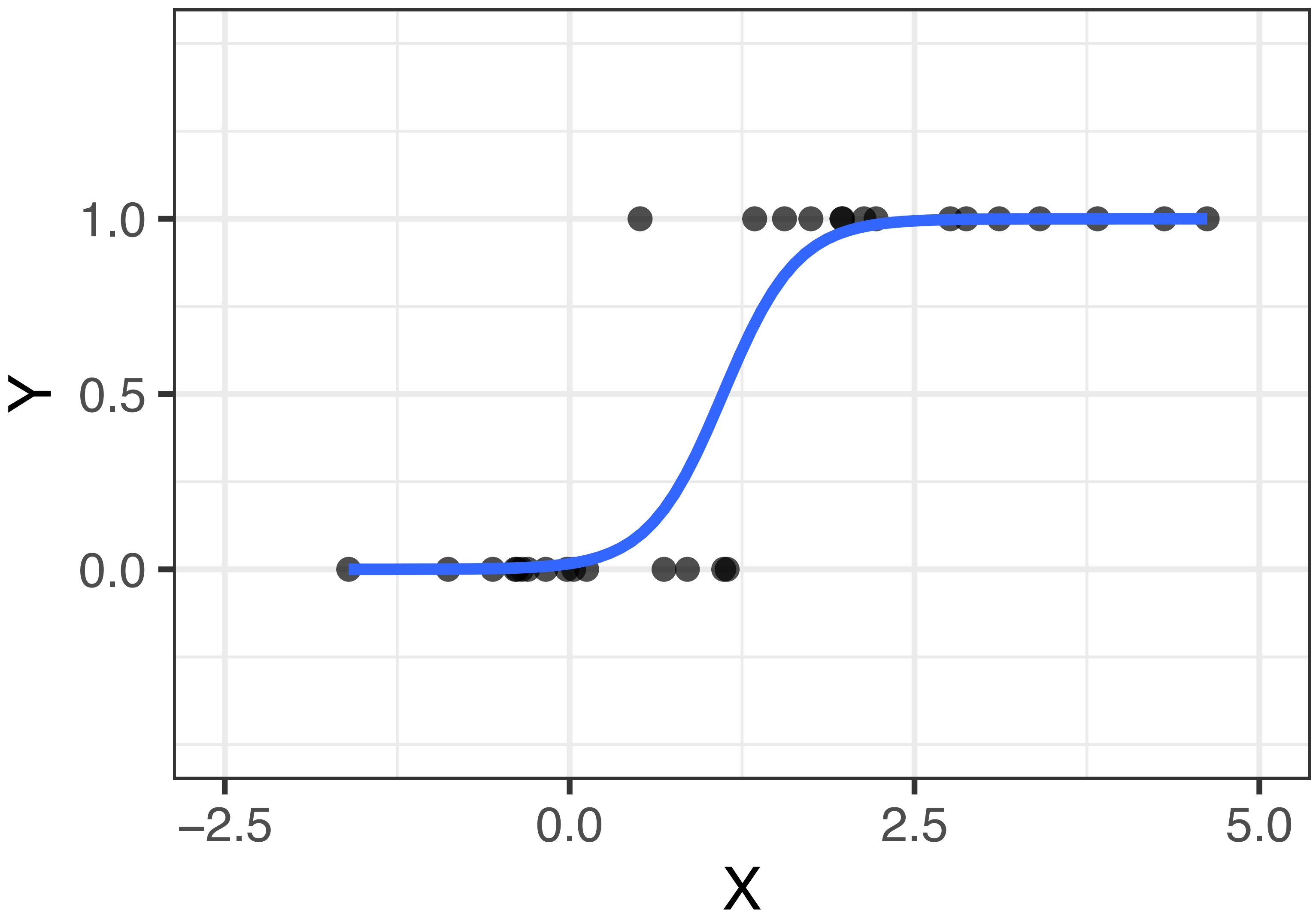
p	odds	log odds
0.2	.25 to 1	-1.39
0.5	1 to 1	0
0.8	4 to 1	1.39

log odds:

$$\ln \frac{p}{1-p}$$

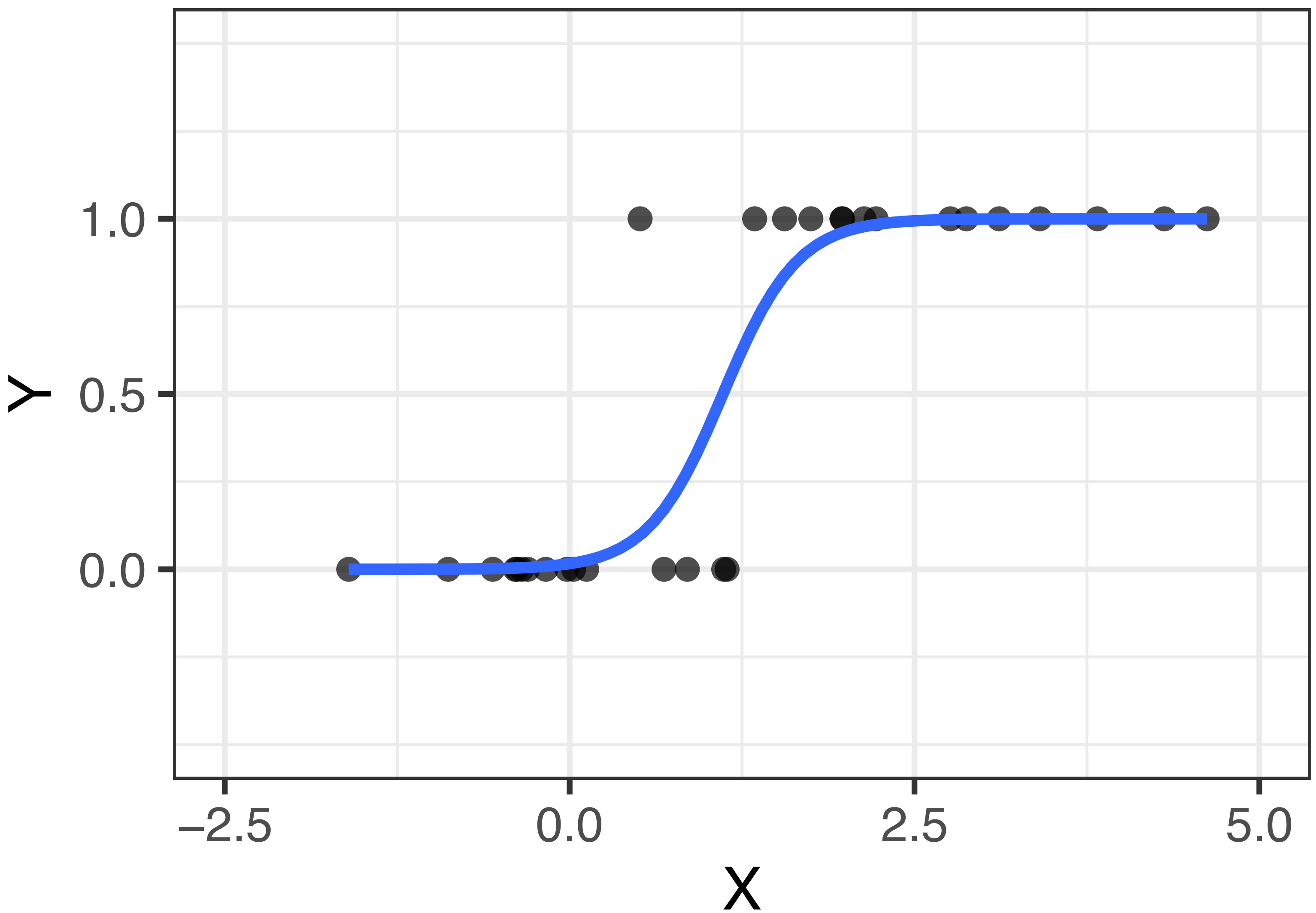
log odds range from -Inf to +Inf

Logistic regression



The logistic function compresses values into [0,1]

Logistic regression

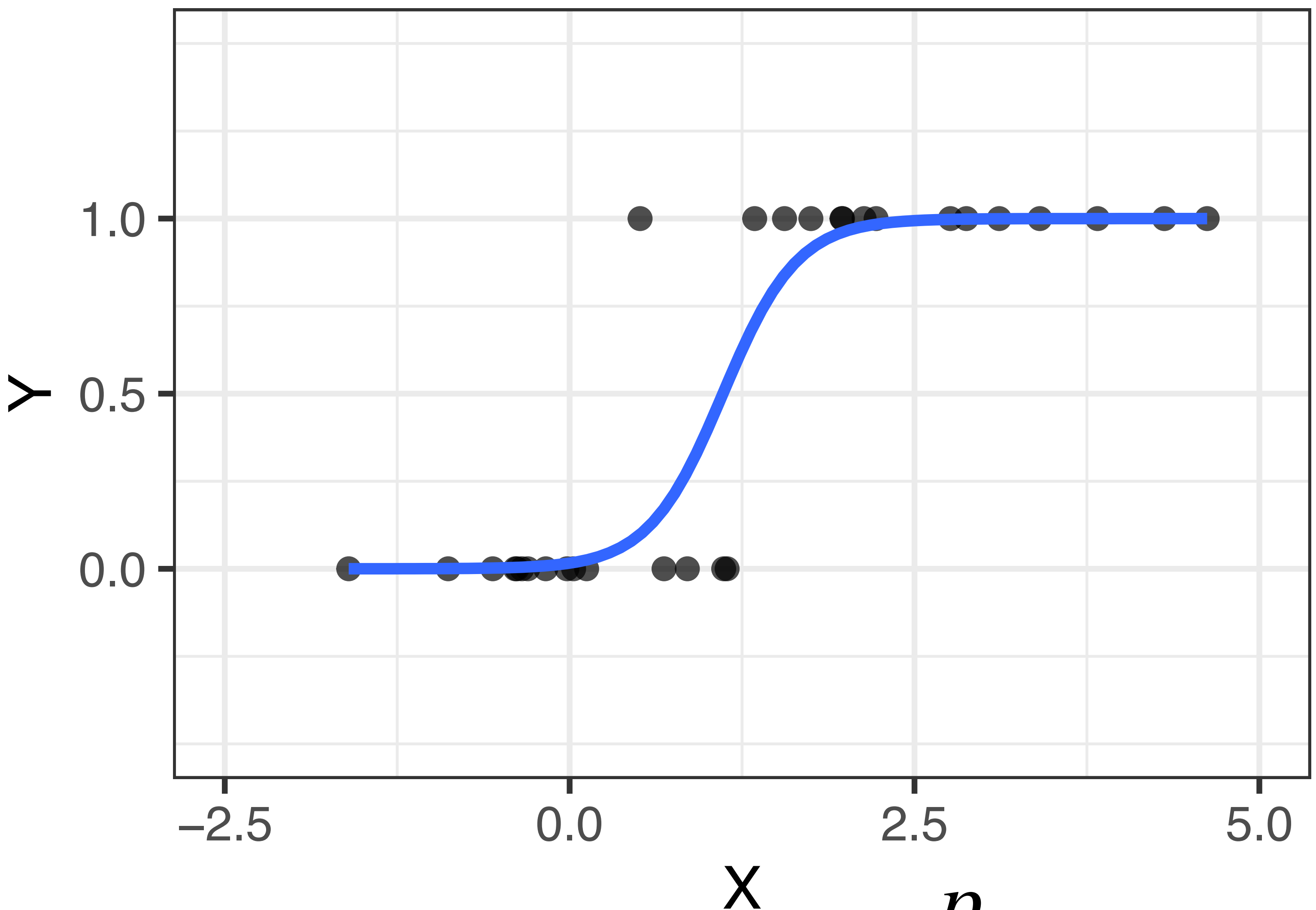


The logistic function compresses values into [0,1]

Logistic regression is a kind of GLM (with binomial link):

$$\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

Logistic regression

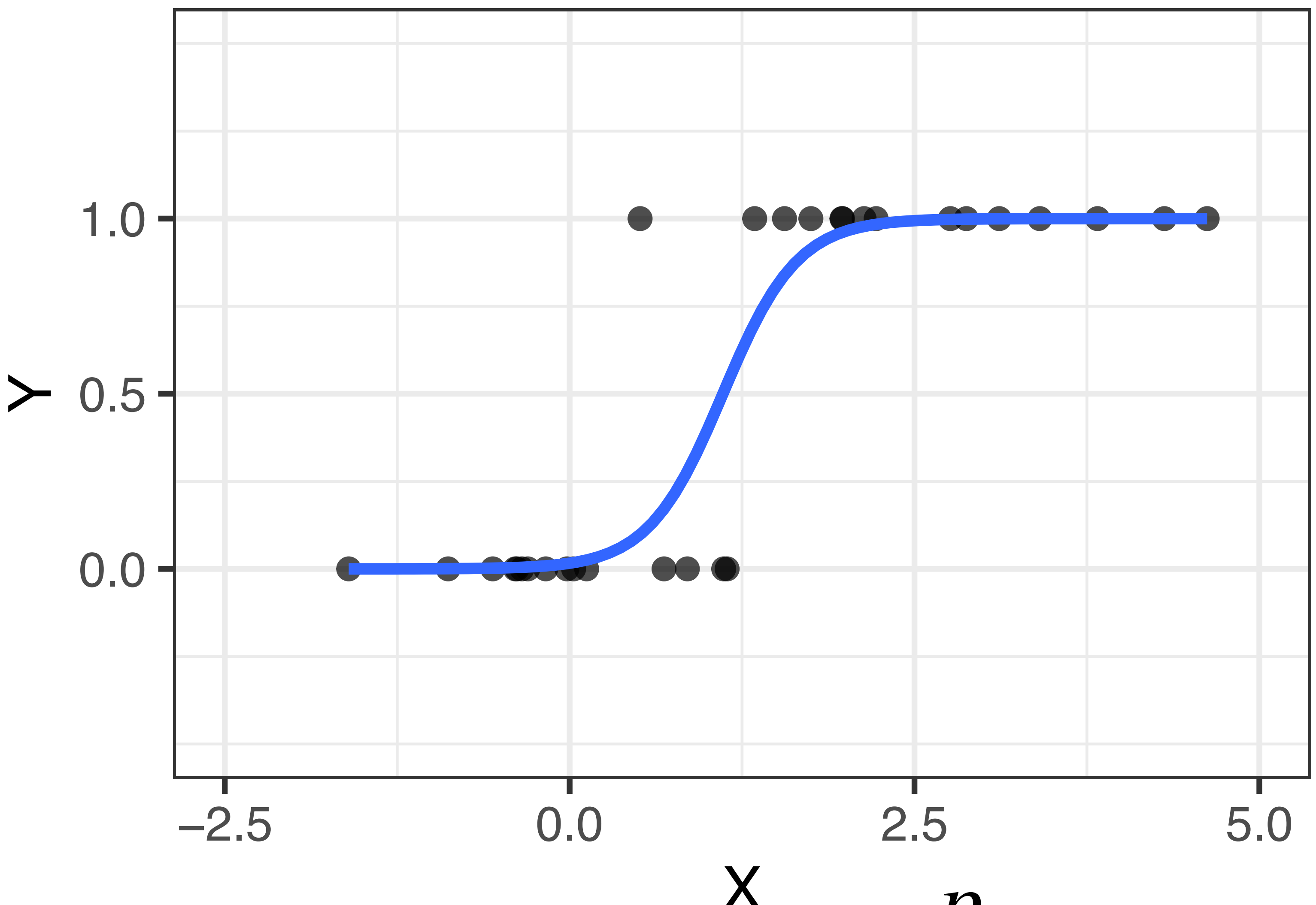


The logistic function compresses values into [0,1]

Logistic regression is a kind of GLM (with binomial link):

$$\ln \frac{p}{1-p} = \eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

Logistic regression

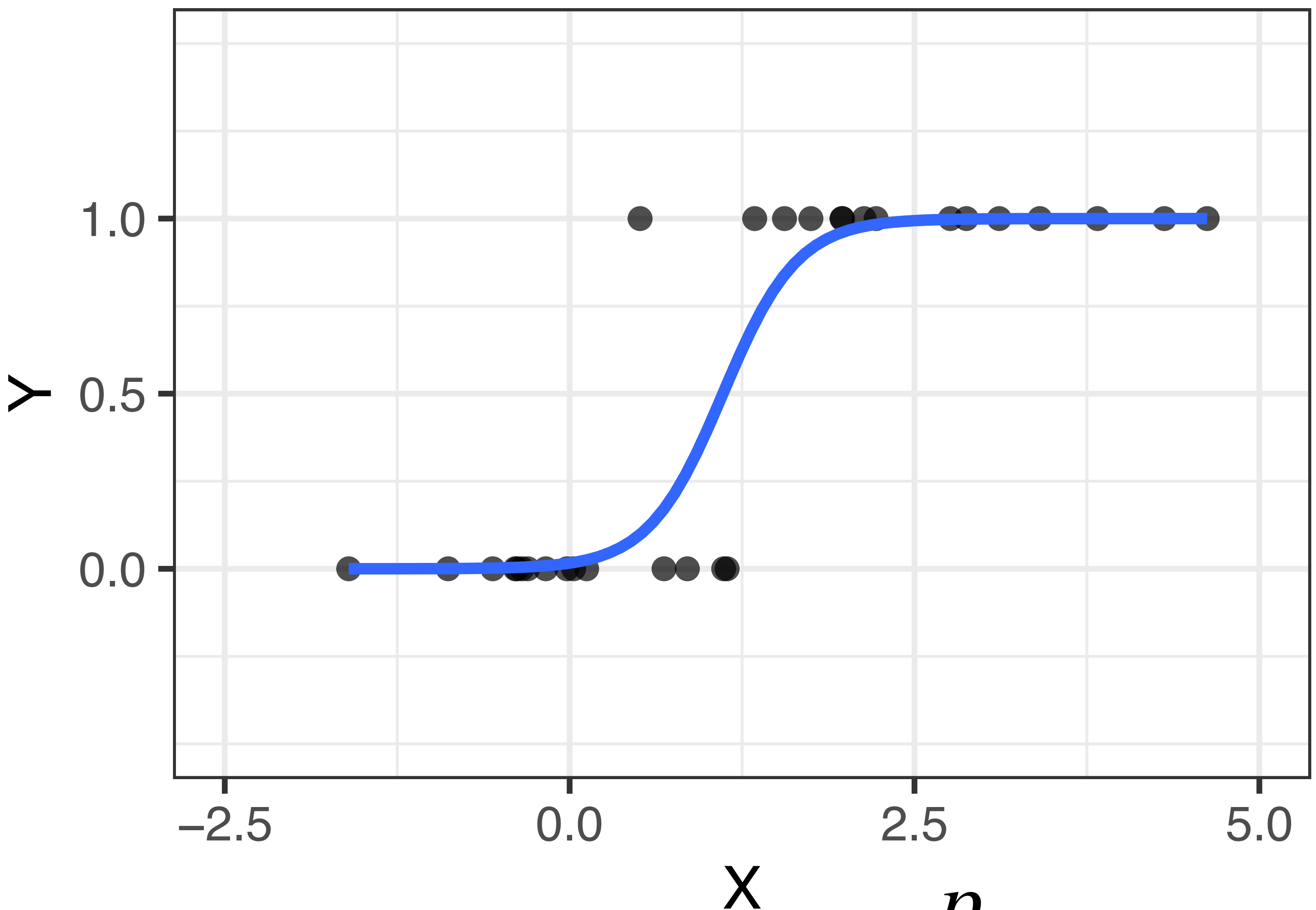


The logistic function compresses values into [0,1]

Logistic regression is a kind of GLM (with binomial link):

$$g(p) = \ln \frac{p}{1-p} = \eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

Logistic regression



The logistic function compresses values into [0,1]

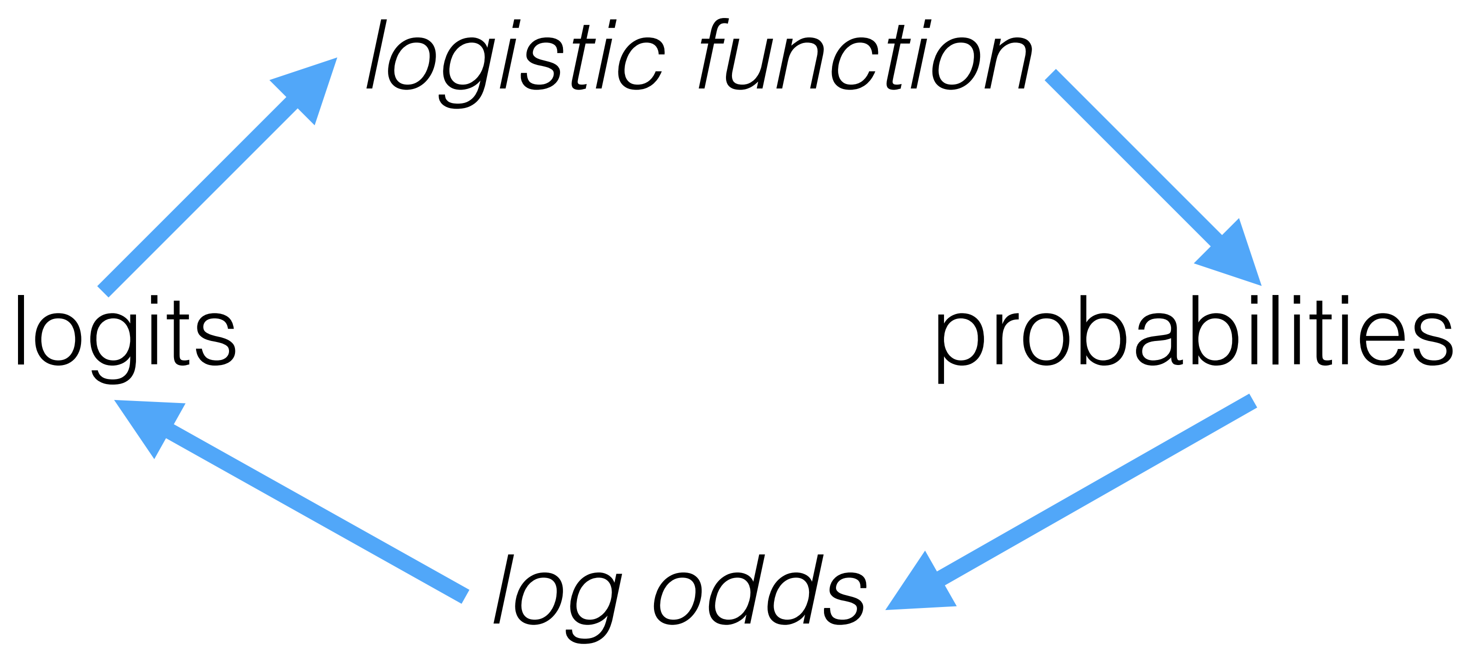
Logistic regression is a kind of GLM (with binomial link):

$$g(p) = \ln \frac{p}{1-p} = \eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

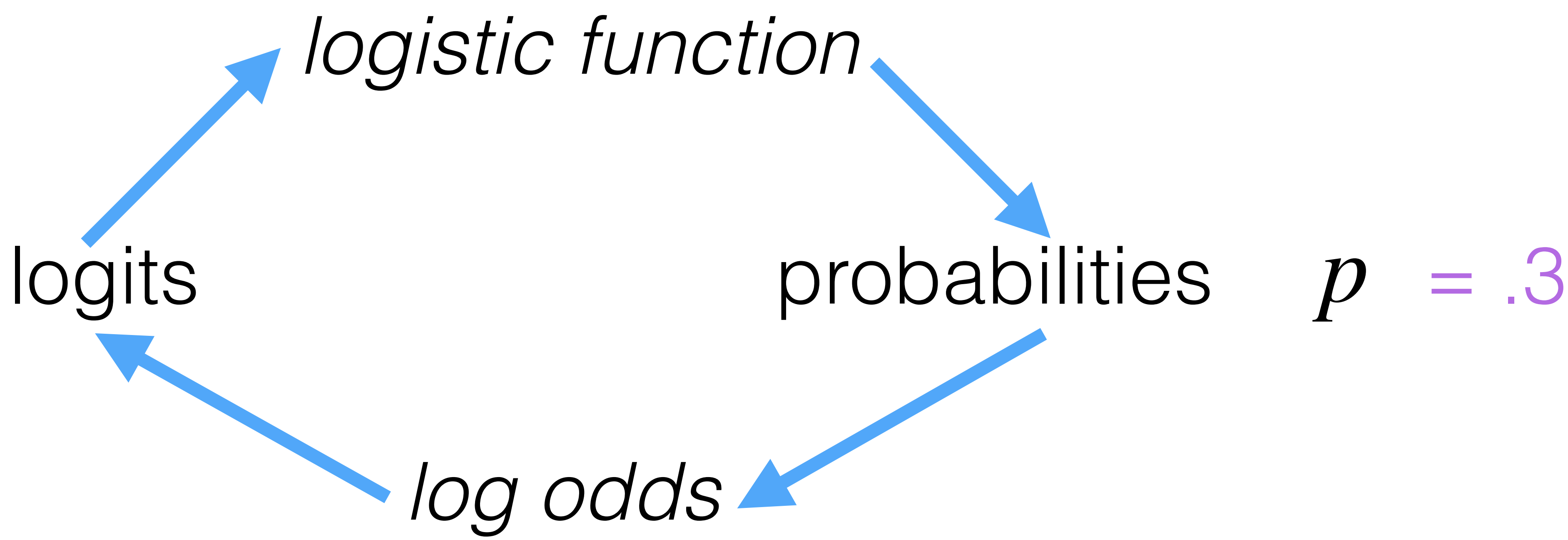
To retrieve p, apply the logistic function (inverse of log odds):

$$p = \frac{1}{1 + \exp^{-X\beta}}$$

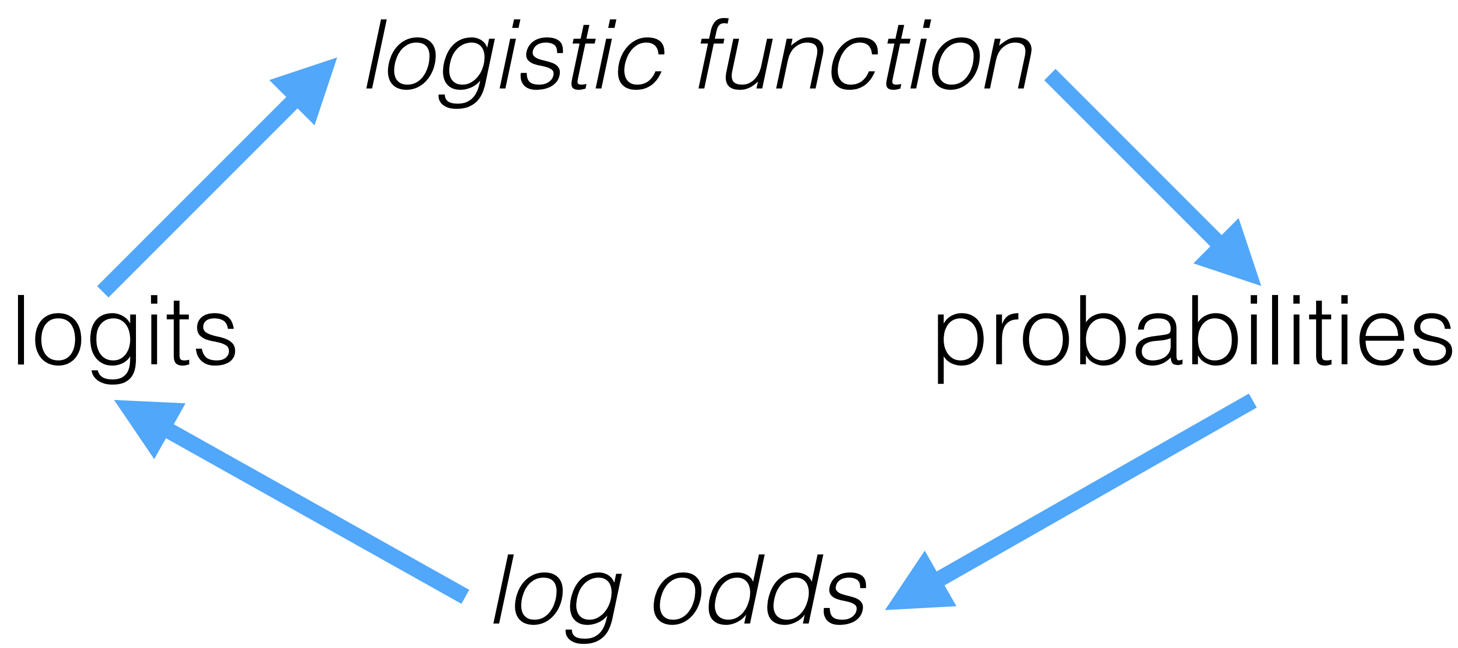
Logistic regression



Logistic regression



Logistic regression



p = .3

$$\text{logit} = \ln \frac{p}{1-p}$$ = ln(.3/.7)

Logistic regression

logistic function

logit logits probabilities p = .3

= -.84

log odds

$$\text{logit} = \ln \frac{p}{1-p}$$ = ln(.3/.7)

Logistic regression

$$p = \frac{1}{1 + \exp^{-\text{logit}}}$$

logistic function

logit = -.84

logits

probabilities

p = .3

log odds

$$\text{logit} = \ln \frac{p}{1-p} = \ln(.3/.7)$$

Mixed effects logistic regression

linear model : mixed linear model ::
logit model : mixed logit model

Assumption: individual differences within a grouping factor are normally distributed in log-odds of event

$$\ln\left(\frac{p}{1-p}\right) = \overbrace{\mathbf{X}\beta}^{\text{Fixed effects}} + \overbrace{\mathbf{Z}b}^{\text{Random effects}} \quad , \quad \overbrace{b_i}^{\sim N(0,\sigma_{b_i})}$$

Dative alternation

Realization of recipient:
NP: John gave [the children] [toys]
PP: John gave [toys] [to the children]

What governs the syntactic choice?

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the Dative Alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), Cognitive Foundations of Interpretation (pp. 1–33). Amsterdam: Royal Netherlands Academy of Science.

Let’s translate it into R!