Do whatever is needed to finish…

EDUC 7610

Chapter 18

Generalized Linear Models (GLM)

Tyson S. Barrett, PhD

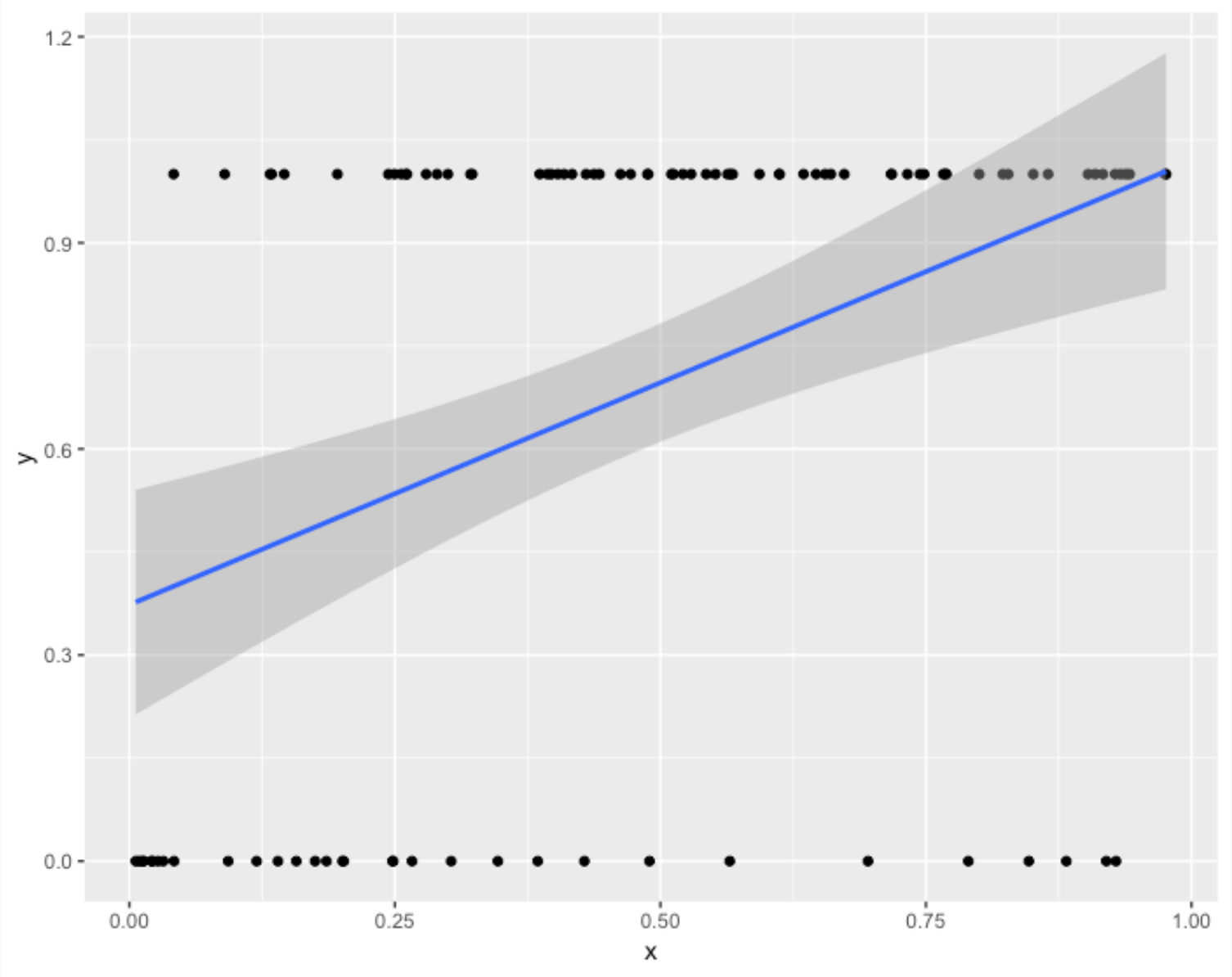
Lots of Types of Outcomes

Not all outcomes are continuous and nicely distributed

Type	Example	Method to Handle It
Dichotomous	Smoker/Non-Smoker Depressed/Not Depressed	Logistic Regression
Count	Number of times visited hospital this month	Poisson Regression, Negative Binomial Regression
Ordinal	Low, Mid, High levels of anxiety	Ordinal Logistic Regression
Time to Event	Time until heart attack	Survival Analysis

What if we just used OLS?

Any issues with this scenario?



The Gist of GLMs

We model the expected value of the model in a different way than regular regression

$$g(Y_i) = \underbrace{\beta_0 + \beta_1 X_1 + \ \ldots + \epsilon_i}_{\text{Predictors (same as in regular regression)}}$$

Link function

Outcome response

The Gist of GLMs

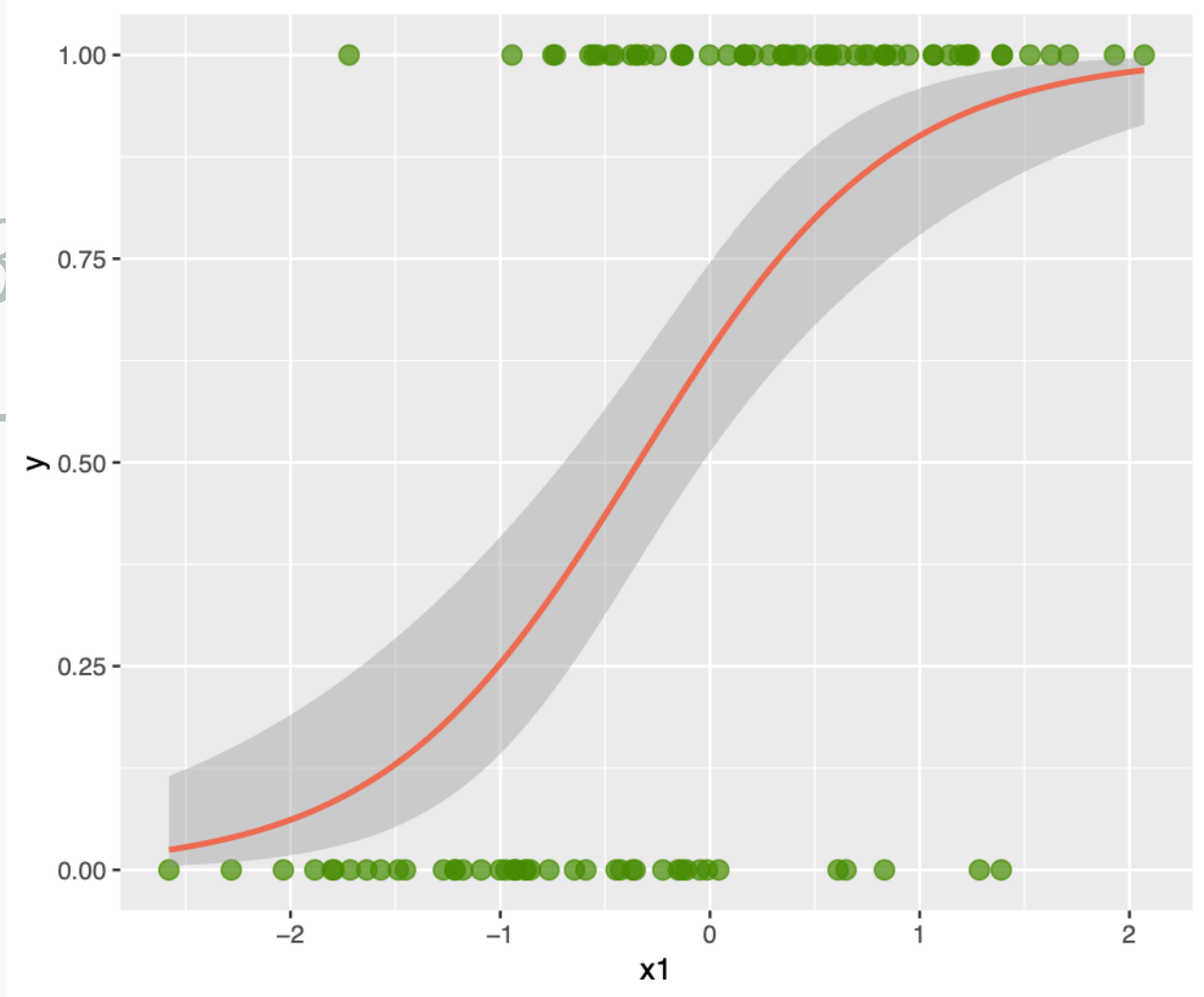
We model the expected value of the model in a different way than regular regression

$$g(Y_i) = \underbrace{\beta_0 + \beta}$$

Link function

(same as

Outcome response



The Gist of GLMs

We model the expected value of the model in a different way than regular regression

$$g(Y_i) = \beta_0 + \beta_1 X_1 + \ \ldots + \epsilon_i$$

Model	Link	Distribution
Linear Regression	Identity	Normal
Logistic Regression	Logit	Binomial
Poisson Regression	Log	Poisson
Loglinear	Log	Poisson
Probit Regression	Probit	Normal

The Linear Models and GLMs

There are so much in common between linear models and generalized linear models

Model specification

Diagnostics

Simple and multiple

Continuous and categorical predictors

Similar assumptions

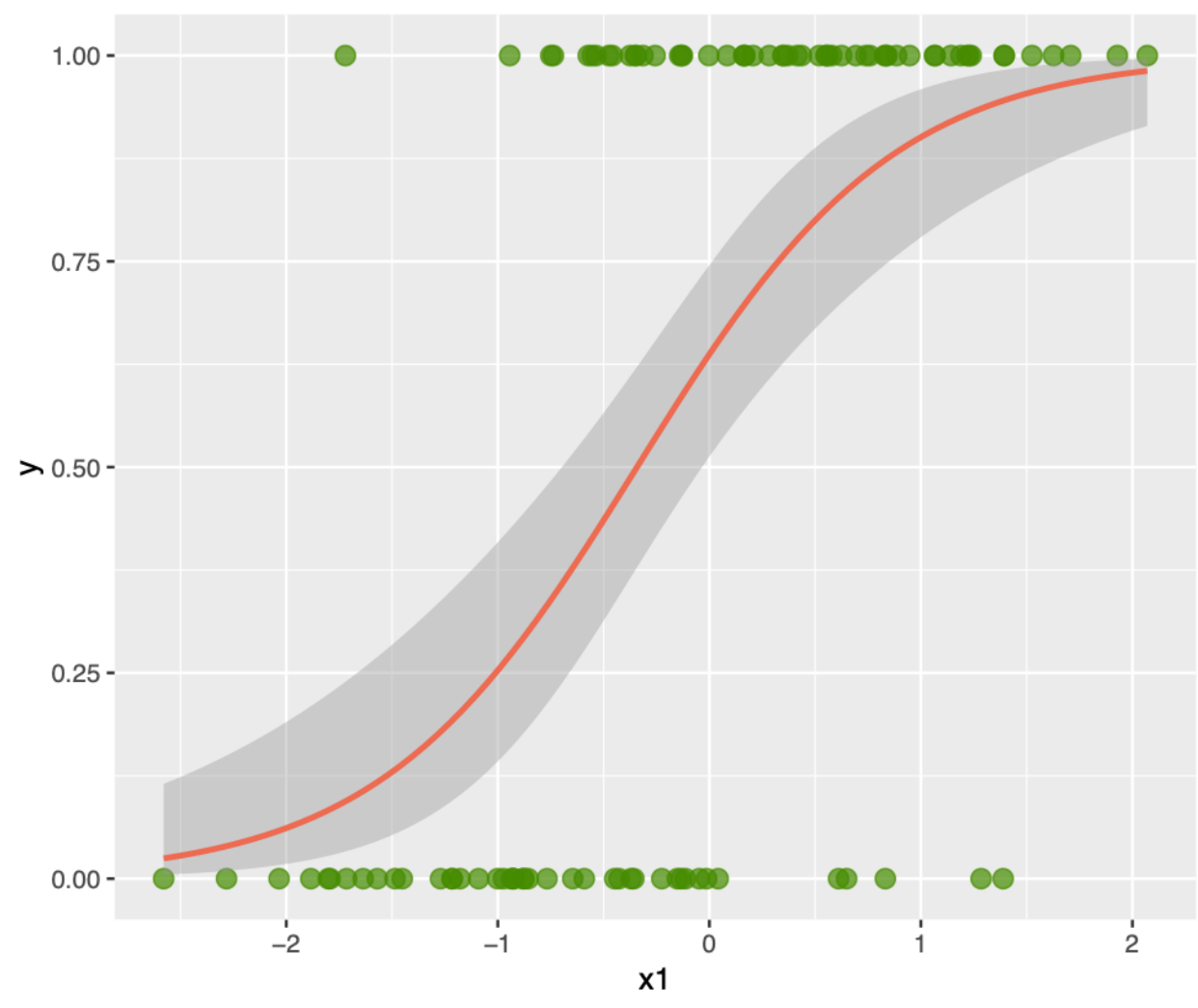
The most common GLM: Logistic

Logistic regression is a particular type of GLM

The outcome is binary (dichotomous)

The predicted values (predicted probabilities) are along an S shaped curve

- Makes it so the predictions are never less than 0 or more than 1
- Often matches how probabilities probably work



The most common GLM: Logistic

Logistic regression is a particular type of GLM

$$\underbrace{logit(Y_i)}_{\log\left(\frac{P}{1-P}\right)} = \beta_0 + \beta_1 X_1 + \ \ldots + \epsilon_i$$

The results then are in terms of "logits" or the "log-odds" of a positive response (gives us the S shape for the predicted values)

For a one unit increase in X_1, there is an associated β_1 change in the log odds of the outcome

Logistic Regression

Since log-odds is not all that intuitive, let's talk about other ways to interpret the results

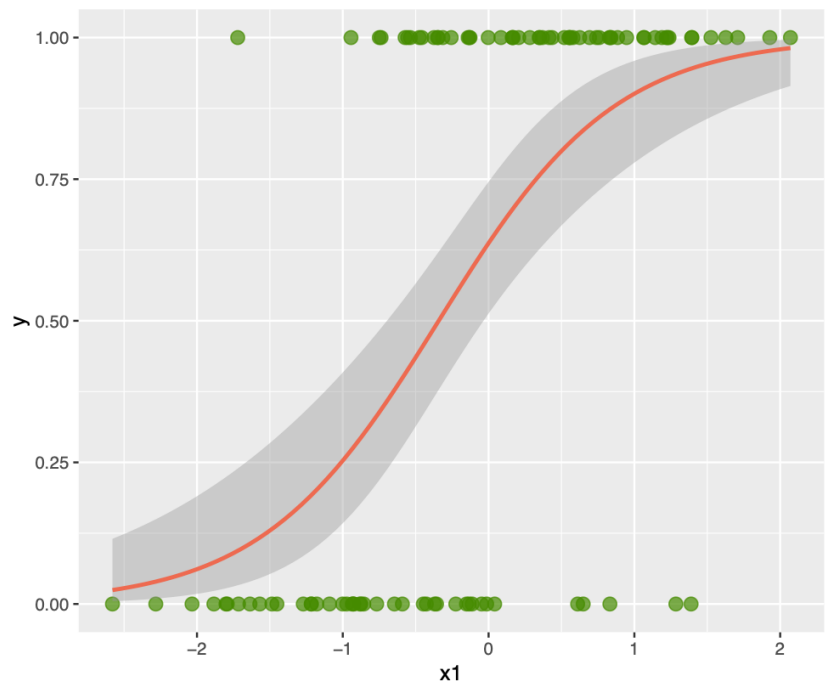
Odds Ratios

Exponentiate the coefficients for OR → e^{β_1}
Can be slightly biased (Mood, 2010) but are the most common way to interpret logistic regression

Average Marginal Effects

Use AMEs to get the average effect in the sample
Are less biased than odds ratios (Mood, 2010)

Predicted Probabilities



Some Additional Linear Models

Poisson Regression

For count outcomes

Ordinal Logistic Regression

For ordinal outcomes

Survival Analysis

For time to event outcomes

Structural Equation Modeling

A flexible, powerful framework for general purpose modeling (linear regression is a subset of SEM)

Time Series

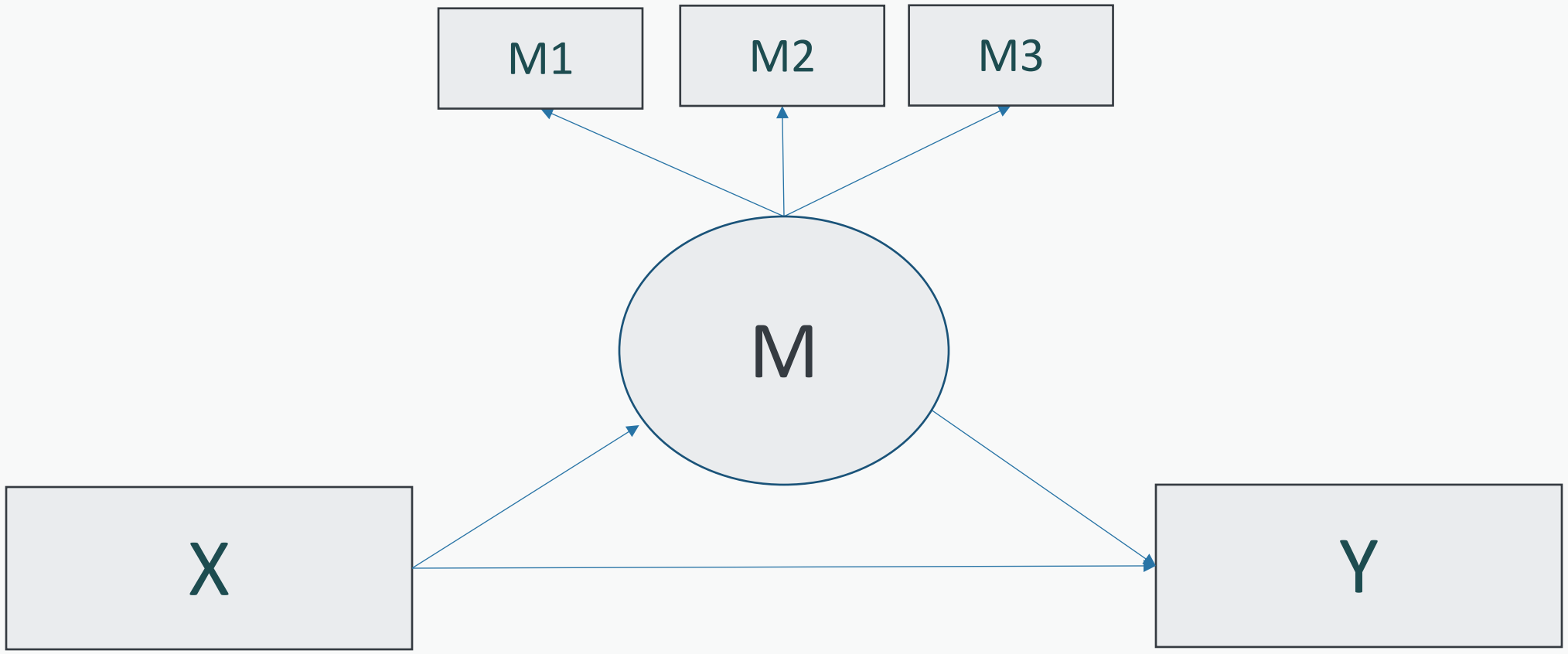
For outcomes where it is measured periodically many times

Multilevel Modeling

For nested or longitudinal outcomes

Structural Equation Modeling

A flexible, powerful framework for general purpose modeling (linear regression is a subset of SEM)



Structural Equation Modeling

A flexible, powerful framework for general purpose modeling (linear regression is a subset of SEM)

1. Can do multiple “dependent” variables
2. Latent variables to control for measurement error
3. Interpreted like regular regression
4. Several approaches (e.g., LCA)
5. Many estimation routines (mostly based on ML like GLMs)
6. More assumptions though