Lesson 26. The Logistic Regression Model

1 Motivation

- Suppose we have:
 - o One binary response variable: 1 or 0, corresponding to yes or no, respectively
 - o One quantitative or categorical explanatory variable
- Visually:



- How can we model this data?
 - The response is yes or no we want to model the probability of success (yes)
 - The range of possible values of a probability is
 - o But... a linear regression model may predict probabilities

- Instead, we can use a <u>transformation</u> that models a <u>curved</u> relationship with horizontal asymptotes at 0 and 1
- Notation:
 - \circ For any fixed value of the predictor x, there are four probabilities:

	True value (population)	Estimate (sample)
"Unconstrained" probability (without model)		
Model-based probability (with model)		

2	Odds						
	• Suppose the probability of success is π						
	• The odds of success is						
	Odds are often expressed as two numbers						
	∘ e.g., "The odds of winning are 2 to 1"						
	• The range of possible values of odds is						
Ex	cample 1.						
	a. If the probability of winning is 0.8, what are the odds of winning?b. If the odds of winning is 0.5, what is the probability of winning?						
3	log(odds)						
3 log(odds)The log(odds) of success is							
- 110 105(0000) 01 0000000 10							
	• The range of possible values of log(odds) is						
Ex	cample 2.						
	a. If the probability of winning is 0.8, what are the log(odds)?b. If the log(odds) of winning is 0.2, what is the probability of winning?						

• The transformation from π to log(odds) is called the **logistic transformation** or **logit transformation**

• Relationships between π , odds(π), and log(odds(π)):

π	$\mathrm{odds}(\pi)$	$\log(\mathrm{odds}(\pi))$
→ 0		
0.5		
→ 1		

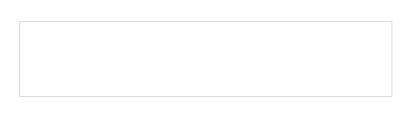
4 The logistic regression model

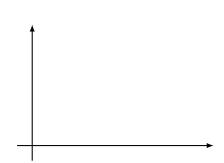
- Variables:
 - o One binary categorical response variable
 - o One explanatory variable, either quantitative or categorical
- Two equivalent forms of the model
- Logit form of model:





• Probability form of model:





Example 3. The dataset MedGPA contains, among other variables, a binary response variable (Acceptance = 1 if accepted, 0 if not), and a quantitative predictor (GPA) for 55 medical school applicants from a college in the Midwest.

- a. Write the logit form of a logistic regression model that predicts the probability of acceptance to medical school based on college GPA.
- b. Write the probability form of the model.

