

Lesson 26. The Logistic Regression Model

1 Motivation

- Suppose we have:
 - One binary response variable: 1 or 0, corresponding to **yes** or **no**, respectively
 - 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
 - But... a linear regression model may predict probabilities
 -
 - Instead, we can use a transformation that models a curved relationship with horizontal asymptotes at 0 and 1
- Notation:
 - 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

Example 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 $\log(\text{odds})$

- The **$\log(\text{odds})$** of success is

- The range of possible values of $\log(\text{odds})$ is

Example 2.

- a. If the probability of winning is 0.8, what are the $\log(\text{odds})$?
- b. If the $\log(\text{odds})$ of winning is 0.2, what is the probability of winning?

- The transformation from π to $\log(\text{odds})$ is called the **logistic transformation** or **logit transformation**

- Relationships between π , $\text{odds}(\pi)$, and $\log(\text{odds}(\pi))$:

π	$\text{odds}(\pi)$	$\log(\text{odds}(\pi))$
$\rightarrow 0$		
0.5		
$\rightarrow 1$		

4 The logistic regression model

- Variables:
 - One binary categorical response variable
 - 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.

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