# Week 4
# Tutorial - Logit Regression

## Applied Statistical Analysis II

Lecturer: Jeffrey Ziegler, PhD
Teaching Fellow: Shekhar Kedia

Spring 2026

## Roadmap for today

- Today:
  - Logistic regression
    1. Estimate model in R
    2. Properly interpret results
    3. Compare different models
- By next week...
  - Work on problem set #2!

# From OLS to Logistic Regression

- Our over-arching goal is still the same: Make inferences about a population from sample data, but now with a binary outcome
- OLS is problematic for 0–1 outcomes
  - Predictions can fall outside $[0, 1]$ and normal, constant-variance errors are implausible
  - Our usual assumptions $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ is not reasonable
    - Distributions of $Y$ don't have same variability across all $x$ values
    - Means of $y$ at different values of $x$ don't have a linear relationship with $x$
- Logistic regression re-expresses the conditional mean $E(Y \mid X)$ as a probability in $(0, 1)$ using a nonlinear link function

# The Logistic Model

- For a binary response $Y_i \in \{0, 1\}$, we model

$$\pi_i = P(Y_i = 1 \mid X_i) = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})}$$

- Equivalently, the *log* odds is linearly associate with our covariates:

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}.$$

- This sets up a generalized linear model estimated by maximum likelihood rather than least squares

# Odds, Odds Ratios, and Interpretation

- The *odds* that $Y = 1$ are:

  $$\text{odds} = \frac{\pi}{1 - \pi},$$

  and the log-odds are:

  $$\text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right).$$

- Interpretation: A one-unit increase in a predictor $X_j$ multiplies the odds by $\exp(\beta_j)$, holding other variables constant
- Ex: If $\exp(\beta_j) = 1.3$, then a one-unit increase in $X_j$ increases the odds of $Y = 1$ by 30%

# Model Assessment and Inference

- Coefficients are tested using Wald tests and confidence intervals based on

$$W = \frac{\hat{\beta}_j}{\mathrm{se}(\hat{\beta}_j)}$$

- Global hypotheses such as

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

  are evaluated using likelihood-ratio $\chi^2$ tests comparing residual deviance from full model to nested model

  - Deviance in logistic regression plays a role analogous to residual sum of squares in OLS, with smaller deviance indicating better fit

# Wrap up: Big Picture

- Logistic regression extends linear modeling to binary outcomes by modeling the log-odds as a linear function of predictors and estimating coefficients via maximum likelihood
- Odds ratios, likelihood-ratio tests, and confidence intervals provide a toolkit for interpretation and model checking similar to what we did last term

# Overview

- Just reviewed how to:
  - Estimate model in R
  - Properly interpret results
  - Compare different models
- Now, let's practice on some real data
- Next week, we'll review what we've learned this term up until now