Binary Outcomes

Francisco Villamil

Applied Quantitative Methods II

MA in Social Sciences, Spring 2026

Today's goals

- Understand why OLS is problematic for binary outcomes

Today's goals

- Understand why OLS is problematic for binary outcomes
- Learn the linear probability model and its trade-offs

Today's goals

- Understand why OLS is problematic for binary outcomes
- Learn the linear probability model and its trade-offs
- Understand logistic regression and how to estimate it in R

Today's goals

- Understand why OLS is problematic for binary outcomes
- Learn the linear probability model and its trade-offs
- Understand logistic regression and how to estimate it in R
- Interpret logit results using marginal effects and predicted probabilities

Today's goals

- Understand why OLS is problematic for binary outcomes
- Learn the linear probability model and its trade-offs
- Understand logistic regression and how to estimate it in R
- Interpret logit results using marginal effects and predicted probabilities
- Compare LPM and logit in practice

Roadmap

Binary outcomes are everywhere

- Many outcomes in social science are binary (yes/no):

Binary outcomes are everywhere

- Many outcomes in social science are binary (yes/no):
  - $\rightarrow$ Did someone vote?

Binary outcomes are everywhere

- Many outcomes in social science are binary (yes/no):
  - $\rightarrow$ Did someone vote?
  - $\rightarrow$ Did a war break out?

Binary outcomes are everywhere

- Many outcomes in social science are binary (yes/no):
  - → Did someone vote?
  - → Did a war break out?
  - → Did a bill pass?

Binary outcomes are everywhere

- Many outcomes in social science are binary (yes/no):
  - → Did someone vote?
  - → Did a war break out?
  - → Did a bill pass?
  - → Did a country democratize?

Binary outcomes are everywhere

- Many outcomes in social science are binary (yes/no):
  - $\rightarrow$ Did someone vote?
  - $\rightarrow$ Did a war break out?
  - $\rightarrow$ Did a bill pass?
  - $\rightarrow$ Did a country democratize?
- Our outcome $Y \in \{0, 1\}$

Binary outcomes are everywhere

- Many outcomes in social science are binary (yes/no):
  - $\rightarrow$ Did someone vote?
  - $\rightarrow$ Did a war break out?
  - $\rightarrow$ Did a bill pass?
  - $\rightarrow$ Did a country democratize?

- Our outcome $Y \in \{0, 1\}$
- We want to model: $\Pr(Y = 1 \mid X)$

What happens if we just use OLS?

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- OLS gives us: $E[Y \mid X] = \beta_0 + \beta_1 X$

What happens if we just use OLS?

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- OLS gives us: $E[Y \mid X] = \beta_0 + \beta_1 X$
- Since $Y \in \{0, 1\}$: $E[Y \mid X] = \Pr(Y = 1 \mid X)$

What happens if we just use OLS?

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- OLS gives us: $E[Y \mid X] = \beta_0 + \beta_1 X$
- Since $Y \in \{0, 1\}$: $E[Y \mid X] = \Pr(Y = 1 \mid X)$

- So OLS is modeling a probability as a **linear function** of X

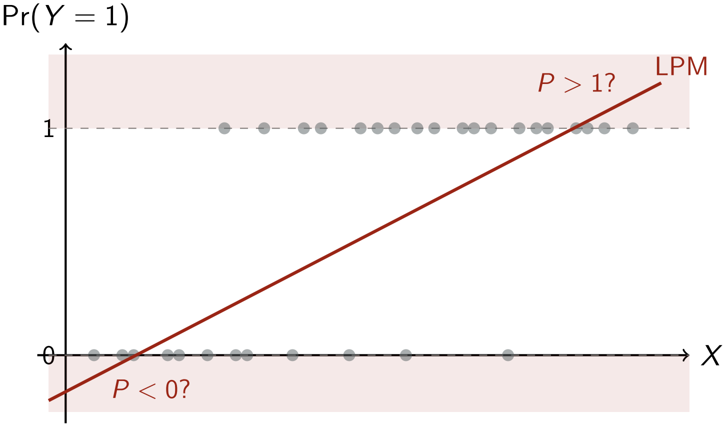
What happens if we just use OLS?

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- OLS gives us: $E[Y \mid X] = \beta_0 + \beta_1 X$
- Since $Y \in \{0, 1\}$: $E[Y \mid X] = \Pr(Y = 1 \mid X)$

- So OLS is modeling a probability as a **linear function** of X
- This is the **linear probability model** (LPM)

The LPM in pictures



The LPM: Simple and intuitive

- β_1 has a direct interpretation:

The LPM: Simple and intuitive

- β_1 has a direct interpretation:
  - $\rightarrow$ A one-unit increase in X changes the probability of $Y = 1$ by β_1

The LPM: Simple and intuitive

- β_1 has a direct interpretation:
  - $\rightarrow$ A one-unit increase in X changes the probability of $Y = 1$ by β_1
- Easy to estimate: just `lm(y ~ x, data = df)`

The LPM: Simple and intuitive

- β_1 has a direct interpretation:
  - $\rightarrow$ A one-unit increase in X changes the probability of $Y = 1$ by β_1
- Easy to estimate: just `lm(y ~ x, data = df)`
- Easy to interpret: same as OLS

The LPM: Simple and intuitive

- β_1 has a direct interpretation:
  - $\rightarrow$ A one-unit increase in X changes the probability of $Y = 1$ by β_1
- Easy to estimate: just `lm(y ~ x, data = df)`
- Easy to interpret: same as OLS
- Many applied researchers use the LPM in practice

The LPM: Simple and intuitive

- β_1 has a direct interpretation:
  - $\rightarrow$ A one-unit increase in X changes the probability of $Y = 1$ by β_1
- Easy to estimate: just `lm(y ~ x, data = df)`
- Easy to interpret: same as OLS
- Many applied researchers use the LPM in practice
- Especially common in economics and political science

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$
  - $\rightarrow$ A linear function can produce $\hat{P} < 0$ or $\hat{P} > 1$

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$
  - $\rightarrow$ A linear function can produce $\hat{P} < 0$ or $\hat{P} > 1$
  - $\rightarrow$ Probabilities must be between 0 and 1

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$
  - $\rightarrow$ A linear function can produce $\hat{P} < 0$ or $\hat{P} > 1$
  - $\rightarrow$ Probabilities must be between 0 and 1
- **Problem 2**: Heteroskedasticity by construction

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$
  - $\rightarrow$ A linear function can produce $\hat{P} < 0$ or $\hat{P} > 1$
  - $\rightarrow$ Probabilities must be between 0 and 1
- **Problem 2**: Heteroskedasticity by construction
  - $\rightarrow$ $\text{Var}(\varepsilon \mid X) = P(1 - P)$, which varies with X

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$
  - $\rightarrow$ A linear function can produce $\hat{P} < 0$ or $\hat{P} > 1$
  - $\rightarrow$ Probabilities must be between 0 and 1
- **Problem 2**: Heteroskedasticity by construction
  - $\rightarrow$ $\text{Var}(\varepsilon \mid X) = P(1 - P)$, which varies with X
  - $\rightarrow$ Always use robust SEs with the LPM

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$
  - $\rightarrow$ A linear function can produce $\hat{P} < 0$ or $\hat{P} > 1$
  - $\rightarrow$ Probabilities must be between 0 and 1
- **Problem 2**: Heteroskedasticity by construction
  - $\rightarrow$ $\text{Var}(\varepsilon \mid X) = P(1 - P)$, which varies with X
  - $\rightarrow$ Always use robust SEs with the LPM
- **Problem 3**: Non-linearity at the extremes

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$
  - $\rightarrow$ A linear function can produce $\hat{P} < 0$ or $\hat{P} > 1$
  - $\rightarrow$ Probabilities must be between 0 and 1
- **Problem 2**: Heteroskedasticity by construction
  - $\rightarrow$ $\text{Var}(\varepsilon \mid X) = P(1 - P)$, which varies with X
  - $\rightarrow$ Always use robust SEs with the LPM
- **Problem 3**: Non-linearity at the extremes
  - $\rightarrow$ True relationship between X and $\Pr(Y = 1)$ is S-shaped

LPM limitations

- **Problem 1**: Predictions outside $[0, 1]$
  - $\rightarrow$ A linear function can produce $\hat{P} < 0$ or $\hat{P} > 1$
  - $\rightarrow$ Probabilities must be between 0 and 1
- **Problem 2**: Heteroskedasticity by construction
  - $\rightarrow$ $\text{Var}(\varepsilon \mid X) = P(1 - P)$, which varies with X
  - $\rightarrow$ Always use robust SEs with the LPM
- **Problem 3**: Non-linearity at the extremes
  - $\rightarrow$ True relationship between X and $\Pr(Y = 1)$ is S-shaped
  - $\rightarrow$ LPM forces it to be linear

When is the LPM “good enough”?

- When probabilities are in the middle range (0.2–0.8)

When is the LPM “good enough”?

- When probabilities are in the middle range (0.2–0.8)
  - $\rightarrow$ The linear approximation is reasonable here

When is the LPM “good enough”?

- When probabilities are in the middle range (0.2–0.8)
  - $\rightarrow$ The linear approximation is reasonable here
- When you care about **average marginal effects**

When is the LPM "good enough"?

- When probabilities are in the middle range (0.2–0.8)
  - $\rightarrow$ The linear approximation is reasonable here
- When you care about **average marginal effects**
  - $\rightarrow$ LPM and logit often give similar AMEs

When is the LPM "good enough"?

- When probabilities are in the middle range (0.2–0.8)
  - $\rightarrow$ The linear approximation is reasonable here
- When you care about **average marginal effects**
  - $\rightarrow$ LPM and logit often give similar AMEs
- When simplicity of interpretation matters

When is the LPM "good enough"?

- When probabilities are in the middle range (0.2–0.8)
  - → The linear approximation is reasonable here
- When you care about **average marginal effects**
  - → LPM and logit often give similar AMEs
- When simplicity of interpretation matters
- When is it **not** good enough?

When is the LPM "good enough"?

- When probabilities are in the middle range (0.2–0.8)
  - $\rightarrow$ The linear approximation is reasonable here
- When you care about **average marginal effects**
  - $\rightarrow$ LPM and logit often give similar AMEs
- When simplicity of interpretation matters
- When is it **not** good enough?
  - $\rightarrow$ Rare events (many observations near 0)

When is the LPM "good enough"?

- When probabilities are in the middle range (0.2–0.8)
  - → The linear approximation is reasonable here
- When you care about **average marginal effects**
  - → LPM and logit often give similar AMEs
- When simplicity of interpretation matters
- When is it **not** good enough?
  - → Rare events (many observations near 0)
  - → You need predicted probabilities to be bounded

When is the LPM "good enough"?

- When probabilities are in the middle range (0.2–0.8)
  - → The linear approximation is reasonable here
- When you care about **average marginal effects**
  - → LPM and logit often give similar AMEs
- When simplicity of interpretation matters
- When is it **not** good enough?
  - → Rare events (many observations near 0)
  - → You need predicted probabilities to be bounded
  - → The relationship is clearly non-linear

You estimate an LPM predicting civil war onset. You find that 8% of predicted probabilities are negative.

Is this a problem? What would you do?

Roadmap

The logistic function

$$\Pr(Y = 1 \mid X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} = \frac{1}{1 + e^{-(\beta_0+\beta_1 X)}}$$

- This is an S-shaped (sigmoid) curve

The logistic function

$$\Pr(Y = 1 \mid X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

- This is an S-shaped (sigmoid) curve
- Output is always between 0 and 1

The logistic function

$$\Pr(Y = 1 \mid X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} = \frac{1}{1 + e^{-(\beta_0+\beta_1 X)}}$$

- This is an S-shaped (sigmoid) curve
- Output is always between 0 and 1
- Steep in the middle, flat at the extremes

The logistic function

$$\Pr(Y = 1 \mid X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

- This is an S-shaped (sigmoid) curve
- Output is always between 0 and 1
- Steep in the middle, flat at the extremes
- A natural model for probabilities

The sigmoid curve

The logit transformation

We can rearrange the logistic function:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X$$

- The left side is the **log-odds** (or "logit")

The logit transformation

We can rearrange the logistic function:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X$$

- The left side is the **log-odds** (or “logit”)
- $\frac{P}{1-P}$ is the **odds** of the event

The logit transformation

We can rearrange the logistic function:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X$$

- The left side is the **log-odds** (or "logit")
- $\frac{P}{1-P}$ is the **odds** of the event
- The model is linear in the log-odds, not in the probability

The logit transformation

We can rearrange the logistic function:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X$$

- The left side is the **log-odds** (or "logit")
- $\frac{P}{1-P}$ is the **odds** of the event
- The model is linear in the log-odds, not in the probability
- This is why we call it "logistic regression" or "logit"

Maximum likelihood estimation

- We can't use OLS for logistic regression

Maximum likelihood estimation

- We can't use OLS for logistic regression
- Instead, we use **maximum likelihood estimation** (MLE)

Maximum likelihood estimation

- We can't use OLS for logistic regression
- Instead, we use **maximum likelihood estimation** (MLE)
- The intuition:

Maximum likelihood estimation

- We can't use OLS for logistic regression
- Instead, we use **maximum likelihood estimation** (MLE)
- The intuition:
  - $\rightarrow$ For each observation, the model predicts $\Pr(Y_i = 1)$

Maximum likelihood estimation

- We can't use OLS for logistic regression
- Instead, we use **maximum likelihood estimation** (MLE)
- The intuition:
  - → For each observation, the model predicts $\Pr(Y_i = 1)$
  - → MLE finds the coefficients that make the observed data most likely

Maximum likelihood estimation

- We can't use OLS for logistic regression
- Instead, we use **maximum likelihood estimation** (MLE)
- The intuition:
  - → For each observation, the model predicts $\Pr(Y_i = 1)$
  - → MLE finds the coefficients that make the observed data most likely
- No need to derive this — R does it for us

Maximum likelihood estimation

- We can't use OLS for logistic regression
- Instead, we use **maximum likelihood estimation** (MLE)
- The intuition:
  - → For each observation, the model predicts $\Pr(Y_i = 1)$
  - → MLE finds the coefficients that make the observed data most likely
- No need to derive this — R does it for us
- The math is different, but the workflow is the same

Estimating logit in R

- The function: `glm(y ~ x, family = binomial, data = df)`

Estimating logit in R

- The function: `glm(y ~ x, family = binomial, data = df)`
- `glm`: generalized linear model

Estimating logit in R

- The function: `glm(y ~ x, family = binomial, data = df)`
- `glm`: generalized linear model
- `family = binomial`: tells R to use logistic regression

Estimating logit in R

- The function: `glm(y ~ x, family = binomial, data = df)`

- `glm`: generalized linear model
- `family = binomial`: tells R to use logistic regression

- The syntax is identical to `lm()`, just swap to `glm()`

Estimating logit in R

- The function: `glm(y ~ x, family = binomial, data = df)`

- `glm`: generalized linear model
- `family = binomial`: tells R to use logistic regression

- The syntax is identical to `lm()`, just swap to `glm()`
- Works with `broom::tidy()`, `modelsummary()`, etc.

Interpreting logit output: Log-odds

- The direct output gives coefficients in **log-odds**

Interpreting logit output: Log-odds

- The direct output gives coefficients in **log-odds**
- $\beta_1 = 0.5$ means:

Interpreting logit output: Log-odds

- The direct output gives coefficients in **log-odds**
- $\beta_1 = 0.5$ means:
  - $\rightarrow$ A one-unit increase in X increases the log-odds by 0.5

Interpreting logit output: Log-odds

- The direct output gives coefficients in **log-odds**
- $\beta_1 = 0.5$ means:
  - $\rightarrow$ A one-unit increase in X increases the log-odds by 0.5
- This is hard to interpret!

Interpreting logit output: Log-odds

- The direct output gives coefficients in **log-odds**
- $\beta_1 = 0.5$ means:
  - $\rightarrow$ A one-unit increase in X increases the log-odds by 0.5
- This is hard to interpret!
- Nobody thinks in log-odds

Interpreting logit output: Odds ratios

- Exponentiate the coefficient: e^{β_1} = odds ratio

Interpreting logit output: Odds ratios

- Exponentiate the coefficient: e^{β_1} = odds ratio
- In R: `exp(coef(model))`

Interpreting logit output: Odds ratios

- Exponentiate the coefficient: e^{β_1} = odds ratio
- In R: `exp(coef(model))`
- $e^{0.5} \approx 1.65$ means:

Interpreting logit output: Odds ratios

- Exponentiate the coefficient: e^{β_1} = odds ratio
- In R: `exp(coef(model))`
- $e^{0.5} \approx 1.65$ means:
  - $\rightarrow$ A one-unit increase in X **multiplies** the odds by 1.65

Interpreting logit output: Odds ratios

- Exponentiate the coefficient: $e^{\beta_1} =$ odds ratio
- In R: `exp(coef(model))`
- $e^{0.5} \approx 1.65$ means:
  - $\rightarrow$ A one-unit increase in X **multiplies** the odds by 1.65
  - $\rightarrow$ Or: the odds increase by 65%

Interpreting logit output: Odds ratios

- Exponentiate the coefficient: $e^{\beta_1} =$ odds ratio
- In R: `exp(coef(model))`
- $e^{0.5} \approx 1.65$ means:
  - $\rightarrow$ A one-unit increase in X **multiplies** the odds by 1.65
  - $\rightarrow$ Or: the odds increase by 65%
- Slightly more intuitive, but still not probabilities

Interpreting logit output: Odds ratios

- Exponentiate the coefficient: e^{β_1} = odds ratio
- In R: `exp(coef(model))`
- $e^{0.5} \approx 1.65$ means:
  - → A one-unit increase in X **multiplies** the odds by 1.65
  - → Or: the odds increase by 65%
- Slightly more intuitive, but still not probabilities
- The change in probability depends on where you start

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$
  - $\rightarrow$ Odds $= e^0 = 1$ (50-50 chance)

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$
  - $\rightarrow$ Odds $= e^0 = 1$ (50-50 chance)
  - $\rightarrow$ Probability $= \frac{1}{1+e^0} = 0.50$

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$
  - $\rightarrow$ Odds $= e^0 = 1$ (50-50 chance)
  - $\rightarrow$ Probability $= \frac{1}{1+e^0} = 0.50$
- For someone with Education $= 8$:

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$
  - $\rightarrow$ Odds $= e^0 = 1$ (50-50 chance)
  - $\rightarrow$ Probability $= \frac{1}{1+e^0} = 0.50$
- For someone with Education $= 8$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 8 = 2$

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$
  - $\rightarrow$ Odds $= e^0 = 1$ (50-50 chance)
  - $\rightarrow$ Probability $= \frac{1}{1+e^0} = 0.50$
- For someone with Education $= 8$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 8 = 2$
  - $\rightarrow$ Odds $= e^2 \approx 7.4$

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$
  - $\rightarrow$ Odds $= e^0 = 1$ (50-50 chance)
  - $\rightarrow$ Probability $= \frac{1}{1+e^0} = 0.50$
- For someone with Education $= 8$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 8 = 2$
  - $\rightarrow$ Odds $= e^2 \approx 7.4$
  - $\rightarrow$ Probability $= \frac{1}{1+e^{-2}} \approx 0.88$

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$
  - $\rightarrow$ Odds $= e^0 = 1 \quad$ (50-50 chance)
  - $\rightarrow$ Probability $= \frac{1}{1+e^0} = 0.50$
- For someone with Education $= 8$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 8 = 2$
  - $\rightarrow$ Odds $= e^2 \approx 7.4$
  - $\rightarrow$ Probability $= \frac{1}{1+e^{-2}} \approx 0.88$
- Going from 4 to 8 years: P goes from 0.50 to 0.88

Worked example: From log-odds to probability

$$\text{Logit model: } \log\left(\frac{P}{1-P}\right) = -2 + 0.5 \cdot \text{Education}$$

- For someone with Education $= 4$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 4 = 0$
  - $\rightarrow$ Odds $= e^0 = 1$ (50-50 chance)
  - $\rightarrow$ Probability $= \frac{1}{1+e^0} = 0.50$
- For someone with Education $= 8$:
  - $\rightarrow$ Log-odds $= -2 + 0.5 \times 8 = 2$
  - $\rightarrow$ Odds $= e^2 \approx 7.4$
  - $\rightarrow$ Probability $= \frac{1}{1+e^{-2}} \approx 0.88$
- Going from 4 to 8 years: P goes from 0.50 to 0.88
- The same 4-unit change gives $\Delta P = 0.38$

Why coefficients alone are not enough

- In OLS: $\beta_1 =$ change in Y for one-unit change in X (always)

Why coefficients alone are not enough

- In OLS: $\beta_1 =$ change in Y for one-unit change in X (always)
- In logit: the change in **probability** depends on:

Why coefficients alone are not enough

- In OLS: $\beta_1 =$ change in Y for one-unit change in X (always)
- In logit: the change in **probability** depends on:
  - $\rightarrow$ The current value of X

Why coefficients alone are not enough

- In OLS: $\beta_1 =$ change in Y for one-unit change in X (always)
- In logit: the change in **probability** depends on:
  - $\rightarrow$ The current value of X
  - $\rightarrow$ The values of all other variables

Why coefficients alone are not enough

- In OLS: $\beta_1 =$ change in Y for one-unit change in X (always)
- In logit: the change in **probability** depends on:
  - $\rightarrow$ The current value of X
  - $\rightarrow$ The values of all other variables
- A coefficient of $\beta_1 = 0.5$ could mean:

Why coefficients alone are not enough

- In OLS: $\beta_1 =$ change in Y for one-unit change in X (always)
- In logit: the change in **probability** depends on:
  - $\rightarrow$ The current value of X
  - $\rightarrow$ The values of all other variables
- A coefficient of $\beta_1 = 0.5$ could mean:
  - $\rightarrow$ Going from $P = 0.01$ to $P = 0.016$ (tiny change)

Why coefficients alone are not enough

- In OLS: $\beta_1 =$ change in Y for one-unit change in X (always)
- In logit: the change in **probability** depends on:
  - $\rightarrow$ The current value of X
  - $\rightarrow$ The values of all other variables
- A coefficient of $\beta_1 = 0.5$ could mean:
  - $\rightarrow$ Going from $P = 0.01$ to $P = 0.016$ (tiny change)
  - $\rightarrow$ Going from $P = 0.50$ to $P = 0.62$ (large change)

Why coefficients alone are not enough

- In OLS: $\beta_1 =$ change in Y for one-unit change in X (always)
- In logit: the change in **probability** depends on:
  - $\rightarrow$ The current value of X
  - $\rightarrow$ The values of all other variables
- A coefficient of $\beta_1 = 0.5$ could mean:
  - $\rightarrow$ Going from $P = 0.01$ to $P = 0.016$ (tiny change)
  - $\rightarrow$ Going from $P = 0.50$ to $P = 0.62$ (large change)
- We need better tools to interpret logit models

Roadmap

A logit model estimates $\hat{\beta}_1 = 0.8$ for education.

Your colleague says: "Education increases the probability of voting by 0.8."

What's wrong with this statement?

Predicted probabilities

- The most intuitive way to interpret logit models

Predicted probabilities

- The most intuitive way to interpret logit models
- "What is the predicted probability of $Y = 1$ for a person with these characteristics?"

Predicted probabilities

- The most intuitive way to interpret logit models
- "What is the predicted probability of $Y = 1$ for a person with these characteristics?"
- In R:

Predicted probabilities

- The most intuitive way to interpret logit models
- "What is the predicted probability of $Y = 1$ for a person with these characteristics?"
- In R:

  $\rightarrow$ `marginaleffects::predictions(model)`

Predicted probabilities

- The most intuitive way to interpret logit models
- "What is the predicted probability of $Y = 1$ for a person with these characteristics?"
- In R:
  - $\rightarrow$ `marginaleffects::predictions(model)`
  - $\rightarrow$ Returns predicted probabilities for each observation

Predicted probabilities

- The most intuitive way to interpret logit models
- "What is the predicted probability of $Y = 1$ for a person with these characteristics?"
- In R:
  - $\rightarrow$ `marginaleffects::predictions(model)`
  - $\rightarrow$ Returns predicted probabilities for each observation
  - $\rightarrow$ Or at specific values: `predictions(model, newdata = datagrid(...))`

Average marginal effects (AME)

- The marginal effect of X on $\Pr(Y = 1)$ varies across observations

Average marginal effects (AME)

- The marginal effect of X on $\Pr(Y = 1)$ varies across observations
- The **AME** averages across all observations in the sample

Average marginal effects (AME)

- The marginal effect of X on $\Pr(Y = 1)$ varies across observations
- The **AME** averages across all observations in the sample
- In R: `marginaleffects::avg_slopes(model)`

Average marginal effects (AME)

- The marginal effect of X on $\Pr(Y = 1)$ varies across observations
- The **AME** averages across all observations in the sample
- In R: `marginaleffects::avg_slopes(model)`
- Interpretation (just like OLS):

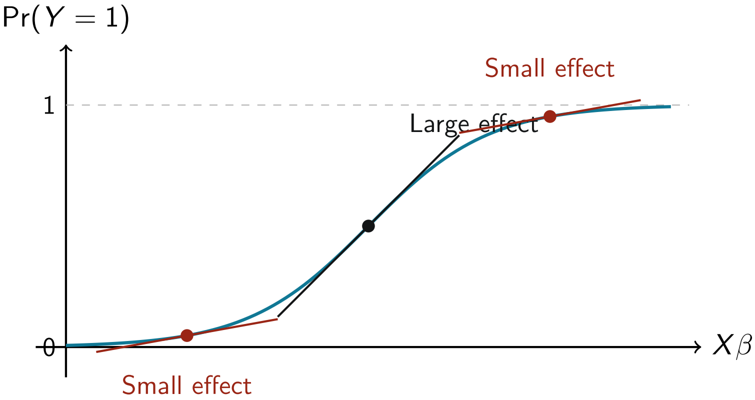
Average marginal effects (AME)

- The marginal effect of X on $\Pr(Y = 1)$ varies across observations
- The **AME** averages across all observations in the sample
- In R: `marginaleffects::avg_slopes(model)`
- Interpretation (just like OLS):
  - $\rightarrow$ "On average, a one-unit increase in X changes $\Pr(Y = 1)$ by ΔP"

Average marginal effects (AME)

- The marginal effect of X on $\Pr(Y = 1)$ varies across observations
- The **AME** averages across all observations in the sample
- In R: `marginaleffects::avg_slopes(model)`
- Interpretation (just like OLS):
  $\rightarrow$ "On average, a one-unit increase in X changes $\Pr(Y = 1)$ by ΔP"
- AMEs from logit are often similar to LPM coefficients

Why marginal effects vary



Marginal effects at representative values

- Instead of averaging, evaluate at specific values

Marginal effects at representative values

- Instead of averaging, evaluate at specific values
- Example: "What is the effect of education on voting for a 40-year-old woman?"

Marginal effects at representative values

- Instead of averaging, evaluate at specific values
- Example: "What is the effect of education on voting for a 40-year-old woman?"
- In R:

Marginal effects at representative values

- Instead of averaging, evaluate at specific values
- Example: “What is the effect of education on voting for a 40-year-old woman?”
- In R:
  - → `avg_slopes(model, newdata = datagrid(age = 40, female = 1))`

Marginal effects at representative values

- Instead of averaging, evaluate at specific values
- Example: "What is the effect of education on voting for a 40-year-old woman?"
- In R:
  - → `avg_slopes(model, newdata = datagrid(age = 40, female = 1))`
- Useful when the effect varies substantially across groups

Marginal effects at representative values

- Instead of averaging, evaluate at specific values
- Example: "What is the effect of education on voting for a 40-year-old woman?"
- In R:
  - → `avg_slopes(model, newdata = datagrid(age = 40, female = 1))`
- Useful when the effect varies substantially across groups
- E.g., the effect may be larger for young people than for old people

Plotting predicted probabilities

- The best way to communicate logit results to any audience

Plotting predicted probabilities

- The best way to communicate logit results to any audience
- Show how $\Pr(Y = 1)$ changes across values of X

Plotting predicted probabilities

- The best way to communicate logit results to any audience
- Show how $\Pr(Y = 1)$ changes across values of X
- Include confidence bands

Plotting predicted probabilities

- The best way to communicate logit results to any audience

- Show how $\Pr(Y = 1)$ changes across values of X
- Include confidence bands

- In R:

Plotting predicted probabilities

- The best way to communicate logit results to any audience
- Show how $\Pr(Y = 1)$ changes across values of X
- Include confidence bands
- In R:
  - $\rightarrow$ `plot_predictions(model, condition = "x")`

Plotting predicted probabilities

- The best way to communicate logit results to any audience
- Show how $\Pr(Y = 1)$ changes across values of X
- Include confidence bands
- In R:
  - → `plot_predictions(model, condition = "x")`
  - → Plots the S-curve with 95% confidence interval

Plotting predicted probabilities

- The best way to communicate logit results to any audience

- Show how $\Pr(Y = 1)$ changes across values of X
- Include confidence bands

- In R:
  - → `plot_predictions(model, condition = "x")`
  - → Plots the S-curve with 95% confidence interval

- Much more informative than a table of log-odds

Comparing LPM and logit

- In many cases, LPM and logit give similar results

Comparing LPM and logit

- In many cases, LPM and logit give similar results
  - $\rightarrow$ Especially for average marginal effects

Comparing LPM and logit

- In many cases, LPM and logit give similar results
  - → Especially for average marginal effects
  - → Especially when probabilities are in the 0.2–0.8 range

Comparing LPM and logit

- In many cases, LPM and logit give similar results
  - → Especially for average marginal effects
  - → Especially when probabilities are in the 0.2–0.8 range
- Where they differ:

Comparing LPM and logit

- In many cases, LPM and logit give similar results
  - → Especially for average marginal effects
  - → Especially when probabilities are in the 0.2–0.8 range
- Where they differ:
  - → Predicted probabilities near 0 or 1

Comparing LPM and logit

- In many cases, LPM and logit give similar results
  - $\rightarrow$ Especially for average marginal effects
  - $\rightarrow$ Especially when probabilities are in the 0.2–0.8 range

- Where they differ:
  - $\rightarrow$ Predicted probabilities near 0 or 1
  - $\rightarrow$ LPM can go outside $[0, 1]$; logit cannot

Comparing LPM and logit

- In many cases, LPM and logit give similar results
  - → Especially for average marginal effects
  - → Especially when probabilities are in the 0.2–0.8 range
- Where they differ:
  - → Predicted probabilities near 0 or 1
  - → LPM can go outside $[0, 1]$; logit cannot
  - → Marginal effects at extreme values

Comparing LPM and logit

- In many cases, LPM and logit give similar results
  - → Especially for average marginal effects
  - → Especially when probabilities are in the 0.2–0.8 range
- Where they differ:
  - → Predicted probabilities near 0 or 1
  - → LPM can go outside $[0, 1]$; logit cannot
  - → Marginal effects at extreme values
- A good practice: estimate both and compare

Model fit for logit

- No R^2 in the usual sense (MLE, not OLS)

Model fit for logit

- No R^2 in the usual sense (MLE, not OLS)
- Alternative measures:

Model fit for logit

- No R^2 in the usual sense (MLE, not OLS)
- Alternative measures:
  - $\rightarrow$ **Pseudo-**R^2: compares model to null model (McFadden's)

Model fit for logit

- No R^2 in the usual sense (MLE, not OLS)
- Alternative measures:
  - $\rightarrow$ **Pseudo-**R^2: compares model to null model (McFadden's)
  - $\rightarrow$ **AIC / BIC**: penalized likelihood (lower $=$ better)

Model fit for logit

- No R^2 in the usual sense (MLE, not OLS)
- Alternative measures:
  - $\rightarrow$ **Pseudo-R^2**: compares model to null model (McFadden's)
  - $\rightarrow$ **AIC / BIC**: penalized likelihood (lower $=$ better)
  - $\rightarrow$ **Classification accuracy**: percent correctly predicted

Model fit for logit

- No R^2 in the usual sense (MLE, not OLS)
- Alternative measures:
  - $\rightarrow$ **Pseudo-**R^2: compares model to null model (McFadden's)
  - $\rightarrow$ **AIC / BIC**: penalized likelihood (lower = better)
  - $\rightarrow$ **Classification accuracy**: percent correctly predicted
  - $\rightarrow$ **ROC / AUC**: trade-off between true and false positives

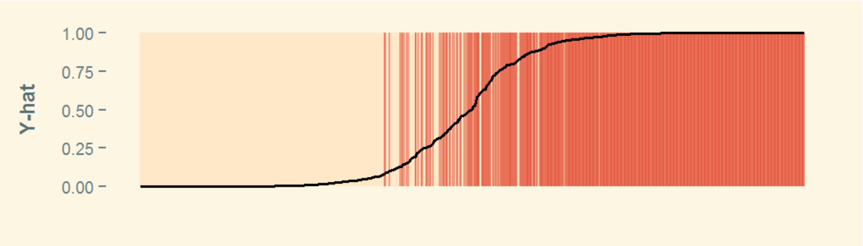
Model fit for logit

- No R^2 in the usual sense (MLE, not OLS)
- Alternative measures:
  - $\rightarrow$ **Pseudo-R^2**: compares model to null model (McFadden's)
  - $\rightarrow$ **AIC / BIC**: penalized likelihood (lower $=$ better)
  - $\rightarrow$ **Classification accuracy**: percent correctly predicted
  - $\rightarrow$ **ROC / AUC**: trade-off between true and false positives
- None is perfect; use as rough guides

Model fit for logit

- No R^2 in the usual sense (MLE, not OLS)
- Alternative measures:
  - $\rightarrow$ **Pseudo-**R^2: compares model to null model (McFadden's)
  - $\rightarrow$ **AIC / BIC**: penalized likelihood (lower = better)
  - $\rightarrow$ **Classification accuracy**: percent correctly predicted
  - $\rightarrow$ **ROC / AUC**: trade-off between true and false positives
- None is perfect; use as rough guides
- Reported automatically by `modelsummary()` and `performance::r2()`

Model fit for rare events: separation plot



- What are we really looking for in a binary model of rare events (e.g. civil wars, genocide, etc)?
- In R: DIY or `separationplot` package

A note on probit

- Logit uses the logistic CDF as the link function

A note on probit

- Logit uses the logistic CDF as the link function
- **Probit** uses the normal (Gaussian) CDF instead

A note on probit

- Logit uses the logistic CDF as the link function
- **Probit** uses the normal (Gaussian) CDF instead
- In R: `glm(y ~ x, family = binomial(link = "probit"), data = df)`

A note on probit

- Logit uses the logistic CDF as the link function
- **Probit** uses the normal (Gaussian) CDF instead
- In R: `glm(y ~ x, family = binomial(link = "probit"), data = df)`
- In practice, logit and probit give very similar results

A note on probit

- Logit uses the logistic CDF as the link function
- **Probit** uses the normal (Gaussian) CDF instead
- In R: `glm(y ~ x, family = binomial(link = "probit"), data = df)`
- In practice, logit and probit give very similar results
  - $\rightarrow$ Probit coefficients $\approx$ logit coefficients $\times 0.625$

A note on probit

- Logit uses the logistic CDF as the link function
- **Probit** uses the normal (Gaussian) CDF instead

- In R: `glm(y ~ x, family = binomial(link = "probit"), data = df)`

- In practice, logit and probit give very similar results
  - $\rightarrow$ Probit coefficients $\approx$ logit coefficients $\times 0.625$
  - $\rightarrow$ Predicted probabilities are nearly identical

A note on probit

- Logit uses the logistic CDF as the link function
- **Probit** uses the normal (Gaussian) CDF instead
- In R: `glm(y ~ x, family = binomial(link = "probit"), data = df)`
- In practice, logit and probit give very similar results
  - → Probit coefficients $\approx$ logit coefficients $\times 0.625$
  - → Predicted probabilities are nearly identical
- Logit is more common in political science; probit in economics

A note on probit

- Logit uses the logistic CDF as the link function
- **Probit** uses the normal (Gaussian) CDF instead
- In R: `glm(y ~ x, family = binomial(link = "probit"), data = df)`
- In practice, logit and probit give very similar results
  - → Probit coefficients ≈ logit coefficients $\times 0.625$
  - → Predicted probabilities are nearly identical
- Logit is more common in political science; probit in economics
- Use whichever is conventional in your field

Roadmap

The Problem with Binary Outcomes

Logistic Regression

Interpreting Logit Results

Practice

Complete workflow in R

```
lpm <- lm(vote ~ age + income + educ, data = df)
logit <- glm(vote ~ age + income + educ,
             family = binomial, data = df)
```

Complete workflow in R

```
lpm <- lm(vote ~ age + income + educ, data = df)
logit <- glm(vote ~ age + income + educ,
             family = binomial, data = df)
modelsummary(list("LPM" = lpm, "Logit" = logit),
             vcov = list("robust", NULL))
```

Complete workflow in R

```
lpm <- lm(vote ~ age + income + educ, data = df)
logit <- glm(vote ~ age + income + educ,
           family = binomial, data = df)
modelsummary(list("LPM" = lpm, "Logit" = logit),
           vcov = list("robust", NULL))
avg_slopes(logit)                    # compare AMEs to LPM
```

Complete workflow in R

```r
lpm <- lm(vote ~ age + income + educ, data = df)
logit <- glm(vote ~ age + income + educ,
           family = binomial, data = df)
modelsummary(list("LPM" = lpm, "Logit" = logit),
           vcov = list("robust", NULL))
avg_slopes(logit)                      # compare AMEs to LPM
plot_predictions(logit, condition = "income")
```

Decision tree: When to use which?

- **Use LPM when:**

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation
  - → Probabilities are in the middle range

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation
  - → Probabilities are in the middle range
  - → You mainly care about average effects

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation
  - → Probabilities are in the middle range
  - → You mainly care about average effects
- **Use logit when:**

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation
  - → Probabilities are in the middle range
  - → You mainly care about average effects
- **Use logit when:**
  - → You need bounded predicted probabilities

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation
  - → Probabilities are in the middle range
  - → You mainly care about average effects
- **Use logit when:**
  - → You need bounded predicted probabilities
  - → Many observations have extreme probabilities (near 0 or 1)

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation
  - → Probabilities are in the middle range
  - → You mainly care about average effects
- **Use logit when:**
  - → You need bounded predicted probabilities
  - → Many observations have extreme probabilities (near 0 or 1)
  - → You want to properly account for the binary nature of Y

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation
  - → Probabilities are in the middle range
  - → You mainly care about average effects
- **Use logit when:**
  - → You need bounded predicted probabilities
  - → Many observations have extreme probabilities (near 0 or 1)
  - → You want to properly account for the binary nature of Y
- In practice: estimate both, report the one most appropriate

Decision tree: When to use which?

- **Use LPM when:**
  - → You want simple, direct interpretation
  - → Probabilities are in the middle range
  - → You mainly care about average effects
- **Use logit when:**
  - → You need bounded predicted probabilities
  - → Many observations have extreme probabilities (near 0 or 1)
  - → You want to properly account for the binary nature of Y
- In practice: estimate both, report the one most appropriate
- Always report marginal effects, not just log-odds

Summary: Key takeaways

- Binary outcomes require special treatment

Summary: Key takeaways

- Binary outcomes require special treatment
- The LPM is simple but has known limitations

Summary: Key takeaways

- Binary outcomes require special treatment
- The LPM is simple but has known limitations
- Logistic regression bounds probabilities between 0 and 1

Summary: Key takeaways

- Binary outcomes require special treatment
- The LPM is simple but has known limitations
- Logistic regression bounds probabilities between 0 and 1
- Log-odds are not intuitive — use marginal effects and predicted probabilities

Summary: Key takeaways

- Binary outcomes require special treatment
- The LPM is simple but has known limitations
- Logistic regression bounds probabilities between 0 and 1
- Log-odds are not intuitive — use marginal effects and predicted probabilities
- AMEs from logit are often similar to LPM coefficients

Summary: Key takeaways

- Binary outcomes require special treatment
- The LPM is simple but has known limitations
- Logistic regression bounds probabilities between 0 and 1
- Log-odds are not intuitive — use marginal effects and predicted probabilities
- AMEs from logit are often similar to LPM coefficients
- Always estimate both and compare; always plot predicted probabilities

For next week

- Read Gelman et al., chapters 11–12
- Read Arel-Bundock, Greifer, and Heiss (2025), chapters 1–4
- Complete Assignment 3

- Next session: Model interpretation and diagnostics
  - → Beyond coefficient tables
  - → Visualizing model results
  - → Residual diagnostics
  - → Influence and outliers

Questions?