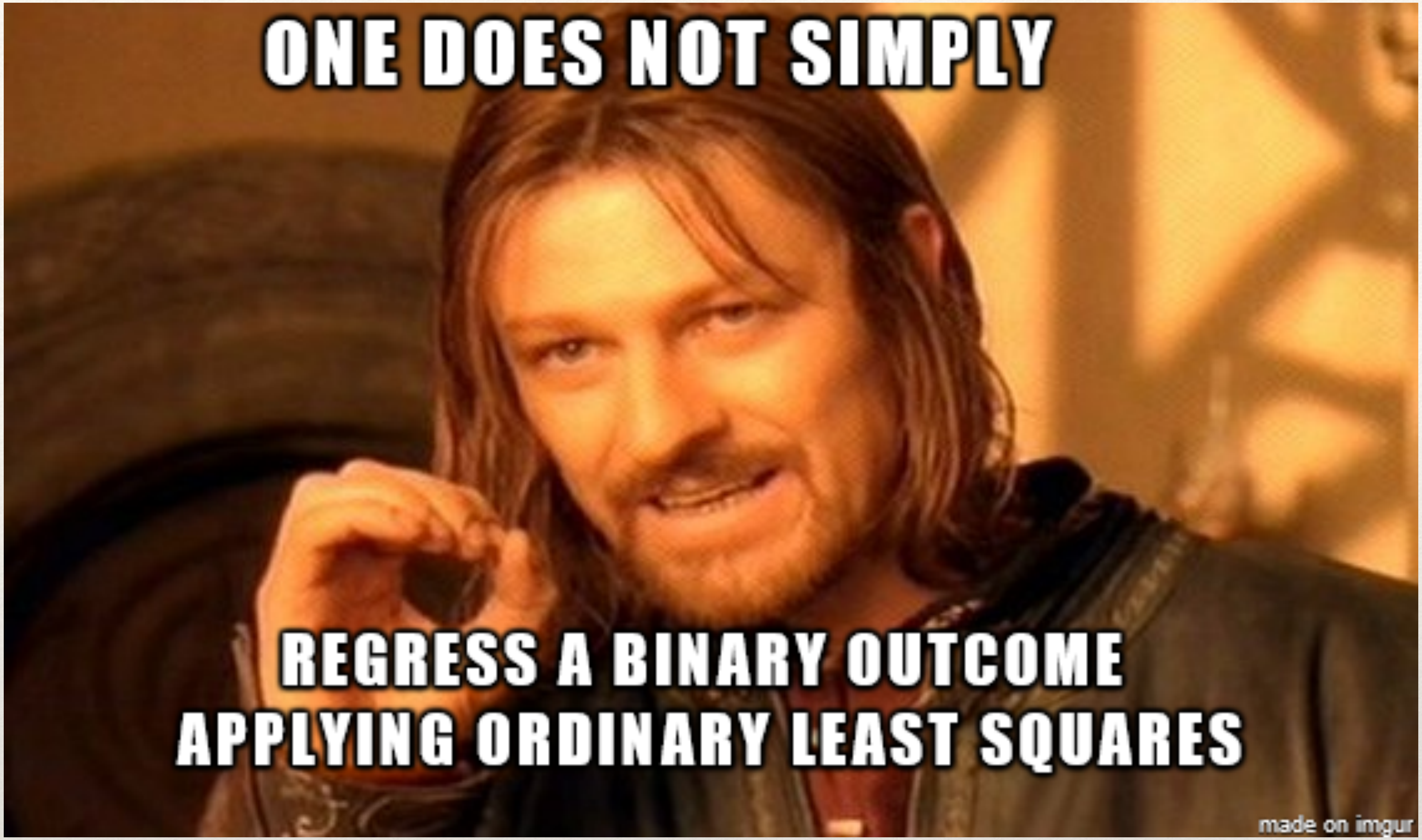


Logistic Regression

Introduction to Statistics

The Plan for Today

The Plan for Today

- **Short Recap:** Linear Regression, Interactions, Polynomials

The Plan for Today

- **Short Recap:** Linear Regression, Interactions, Polynomials
- **Logistic Regression**

The Plan for Today

- **Short Recap:** Linear Regression, Interactions, Polynomials
- **Logistic Regression**
  - Predicting discrete binary outcomes: **Elected**/Not Elected, **War**/Not War, **Voted**/Didn't Vote, **Voted Labour**/Did not vote Labour

The Plan for Today

- **Short Recap:** Linear Regression, Interactions, Polynomials
- **Logistic Regression**
  - Predicting discrete binary outcomes: **Elected**/Not Elected, **War**/Not War, **Voted**/Didn't Vote, **Voted Labour**/Did not vote Labour
  - Minimal aims: summarising + visualising relationships.

The Plan for Today

- **Short Recap:** Linear Regression, Interactions, Polynomials
- **Logistic Regression**
  - Predicting discrete binary outcomes: **Elected**/Not Elected, **War**/Not War, **Voted**/Didn't Vote, **Voted Labour**/Did not vote Labour
  - Minimal aims: summarising + visualising relationships.
  - Going a bit further: understanding **average marginal effects**, very gentle introduction to **maximum likelihood estimation**.

The Plan for Today

- **Short Recap:** Linear Regression, Interactions, Polynomials
- **Logistic Regression**
  - Predicting discrete binary outcomes: **Elected**/Not Elected, **War**/Not War, **Voted**/Didn't Vote, **Voted Labour**/Did not vote Labour
  - Minimal aims: summarising + visualising relationships.
  - Going a bit further: understanding **average marginal effects**, very gentle introduction to **maximum likelihood estimation**.
- **Course Wrap-up**

The Plan for Today

- **Short Recap:** Linear Regression, Interactions, Polynomials
- **Logistic Regression**
  - Predicting discrete binary outcomes: **Elected**/Not Elected, **War**/Not War, **Voted**/Didn't Vote, **Voted Labour**/Did not vote Labour
  - Minimal aims: summarising + visualising relationships.
  - Going a bit further: understanding **average marginal effects**, very gentle introduction to **maximum likelihood estimation**.
- **Course Wrap-up**
  - Where to go next, Q&A.

Recap: Multiple Linear Regression

Recap: Multiple Linear Regression

- An outcome variable Y is generated in the population as a linear combination of variables plus some chance error ϵ:

Recap: Multiple Linear Regression

- An outcome variable Y is generated in the population as a linear combination of variables plus some chance error ϵ:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

Recap: Multiple Linear Regression

- An outcome variable Y is generated in the population as a linear combination of variables plus some chance error ϵ:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

Recap: Multiple Linear Regression

- An outcome variable Y is generated in the population as a linear combination of variables plus some chance error ϵ:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

- Our data are a sample from this population.

Recap: Multiple Linear Regression

- An outcome variable Y is generated in the population as a linear combination of variables plus some chance error ϵ:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

- Our data are a sample from this population.
- We estimate $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3 \ldots \hat{\beta}_p$ so that the sum of squared residuals (the errors we observed in the sample) is minimised.

Recap: Multiple Linear Regression

- An outcome variable Y is generated in the population as a linear combination of variables plus some chance error ϵ:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_p X_p + \epsilon$$

- Our data are a sample from this population.
- We estimate $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3 \ldots \hat{\beta}_p$ so that the sum of squared residuals (the errors we observed in the sample) is minimised.
- This procedure recovers the population parameters *without bias* and *efficiently* under some strong assumptions about model specification and the nature of the error term.

Recap: Multiple Linear Regression

Recap: Multiple Linear Regression

* The coefficients returned by a multiple linear regression represent the expected change in *Y* associated with a **one-unit increase** in *X*, holding all other covariates constant.

Recap: Multiple Linear Regression

- The coefficients returned by a multiple linear regression represent the expected change in Y associated with a **one-unit increase** in X, holding all other covariates constant.
- When a variable is nominal, each category will have its own coefficient, which refers to the **expected difference** in the outcome between that category and the 'reference group'.

Recap: Multiple Linear Regression

* The coefficients returned by a multiple linear regression represent the expected change in Y associated with a **one-unit increase** in X, holding all other covariates constant.
* When a variable is nominal, each category will have its own coefficient, which refers to the **expected difference** in the outcome between that category and the 'reference group'.
* Standard errors represent the **uncertainty** of the coefficient estimate. P-value summarise our evidence against the null that the coefficient is zero in the population.

Recap: Interactions and Polynomials

Recap: Interactions and Polynomials

* Interactions are used to model **moderation**: when the relationship between X_1 and Y depends on the level of a third variable X_2.

Recap: Interactions and Polynomials

* Interactions are used to model **moderation**: when the relationship between X_1 and Y depends on the level of a third variable X_2.
* When X_1 is our independent variable of interest, we call this **heterogeneous treatment effect**. Modelled as:

Recap: Interactions and Polynomials

* Interactions are used to model **moderation**: when the relationship between X_1 and Y depends on the level of a third variable X_2.
* When X_1 is our independent variable of interest, we call this **heterogeneous treatment effect**. Modelled as:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \cdot X_2) + \epsilon$$

Recap: Interactions and Polynomials

- Interactions are used to model **moderation**: when the relationship between X_1 and Y depends on the level of a third variable X_2.
- When X_1 is our independent variable of interest, we call this **heterogeneous treatment effect**. Modelled as:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \cdot X_2) + \epsilon$$

- The effect of X_1 linearly depends on X_2. As we increase X_2 by one unit, the effect of a one-unit increase of X_1 on Y goes up by β_3.

Recap: Interactions and Polynomials

Recap: Interactions and Polynomials

- Polynomials are used to model **non-linearity**: when the relationship between X and Y depends on the level of X itself.

Recap: Interactions and Polynomials

- Polynomials are used to model **non-linearity**: when the relationship between X and Y depends on the level of X itself.
- Most commonly: second-order (quadratic) polynomial:

Recap: Interactions and Polynomials

- Polynomials are used to model **non-linearity**: when the relationship between X and Y depends on the level of X itself.
- Most commonly: second-order (quadratic) polynomial:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \epsilon$$

Recap: Interactions and Polynomials

* Polynomials are used to model **non-linearity**: when the relationship between X and Y depends on the level of X itself.
* Most commonly: second-order (quadratic) polynomial:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \epsilon$$

* Graphically, a parabola with vertex at $-\beta_1/2\beta_2$. U-shaped if $\beta_2 > 0$, n-shaped if $\beta_2 < 0$.

Recap: Interactions and Polynomials

- Polynomials are used to model **non-linearity**: when the relationship between X and Y depends on the level of X itself.
- Most commonly: second-order (quadratic) polynomial:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \epsilon$$

- Graphically, a parabola with vertex at $-\beta_1 / 2\beta_2$. U-shaped if $\beta_2 > 0$, n-shaped if $\beta_2 < 0$.
- Slope varies across values of X: instantaneous rate of change is $\beta_1 + 2\beta_2 X$. (The derivative, which will come back today!)

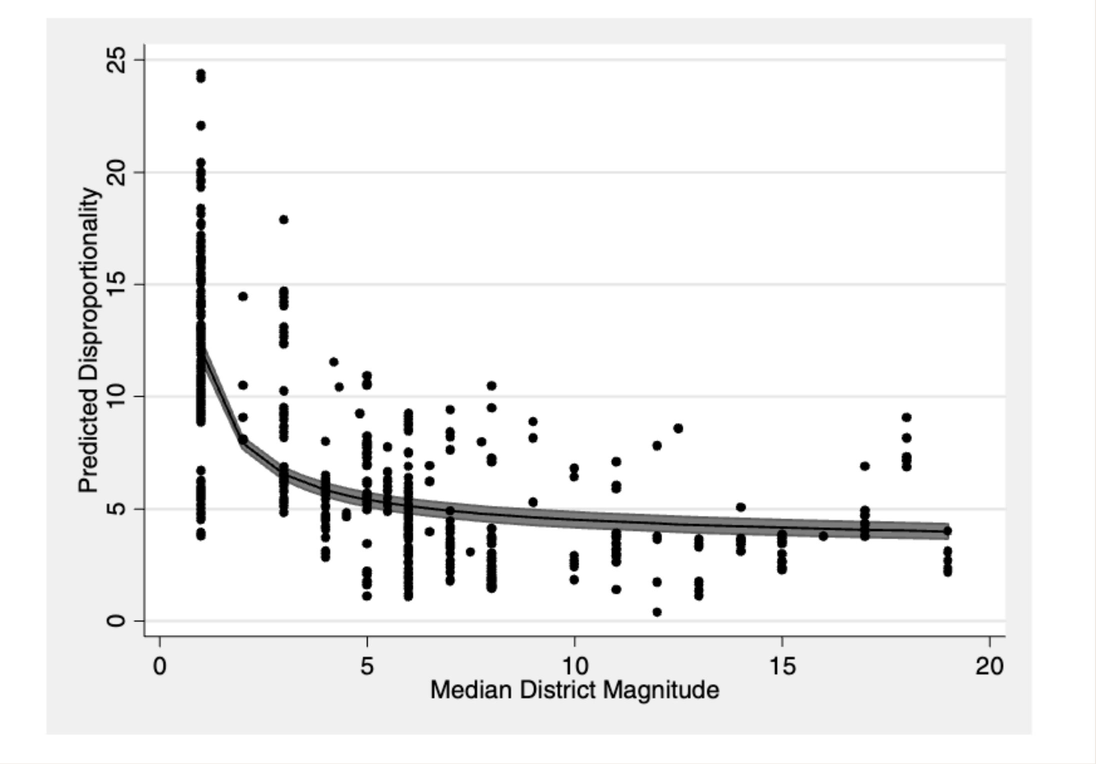
Recap: Visualising Model Estimates

Recap: Visualising Model Estimates

* As our models get more complex, tables < plots.

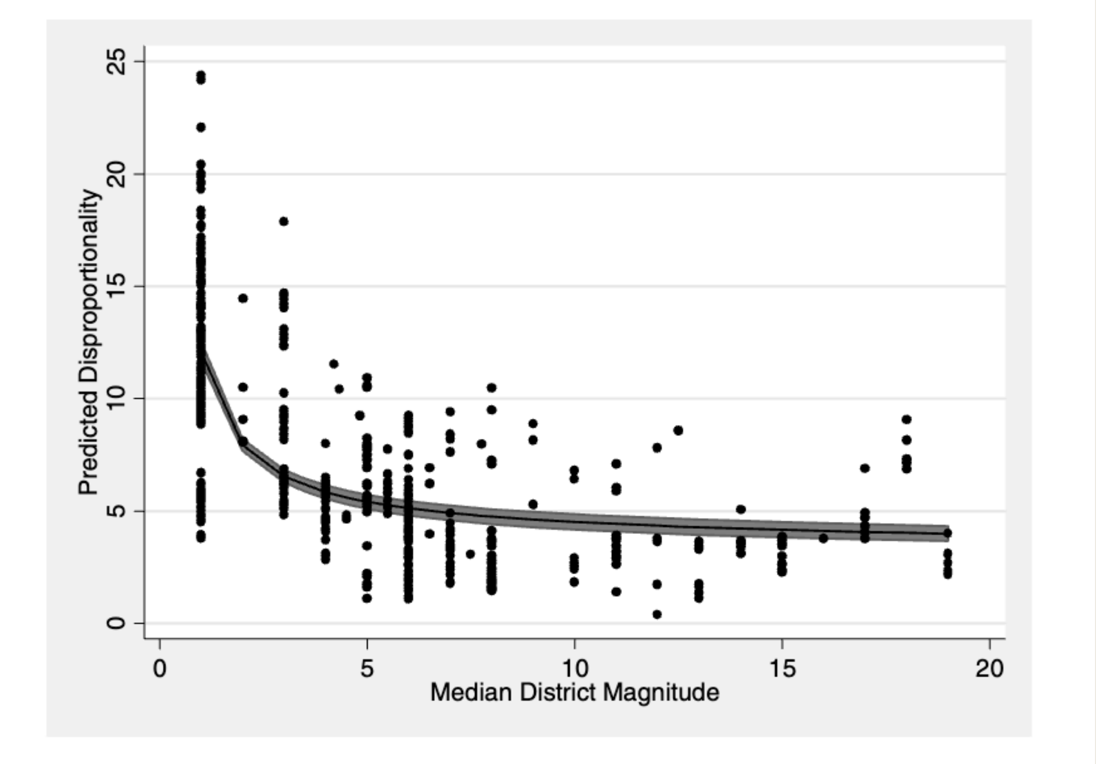
Recap: Visualising Model Estimates

- As our models get more complex, tables < plots.
- **Predicted Values Plot**. Plot the predicted values of *Y* across values of *X*, holding controls constant.



Recap: Visualising Model Estimates

- As our models get more complex, tables < plots.
- **Predicted Values Plot**. Plot the predicted values of *Y* across values of *X*, holding controls constant.
- **Conditional Effect Plots**. Plot the marginal effect of *X* on *Y* across values of *Z* (moderation) or *X* itself (non-linearity).



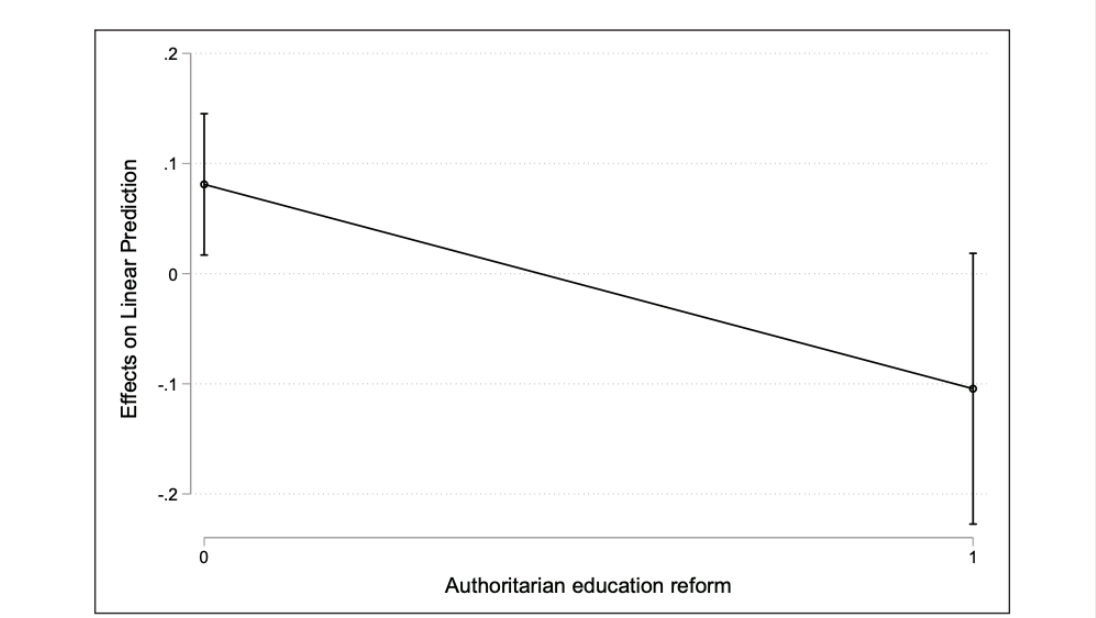


Figure 2. Marginal Effects Plot: Differentiating the Effect of Democratic and Authoritarian Reforms on Satisfaction With Democracy.
Estimates from Model 5 in Table 3.

Logistic Regression

Logistic Regression: Motivation

Logistic Regression: Motivation

- What does it mean predicting ‘turnout’, or ‘war’ or ‘Leave vote’?

Logistic Regression: Motivation

* What does it mean predicting ‘turnout’, or ‘war’ or ‘Leave vote’?
* When our dependent variable is binary, we would like to know the **probability** of an outcome (vote, war, Leave vote) from predictors.

Logistic Regression: Motivation

* What does it mean predicting ‘turnout’, or ‘war’ or ‘Leave vote’?
* When our dependent variable is binary, we would like to know the **probability** of an outcome (vote, war, Leave vote) from predictors.
* An approach: just use OLS.

Logistic Regression: Motivation

- What does it mean predicting ‘turnout’, or ‘war’ or ‘Leave vote’?
- When our dependent variable is binary, we would like to know the **probability** of an outcome (vote, war, Leave vote) from predictors.
- An approach: just use OLS.

$$\Pr(\text{Leave}_i = 1) = \alpha + \beta\text{Euroscepticism (0-10 scale)}_i + \epsilon_i$$

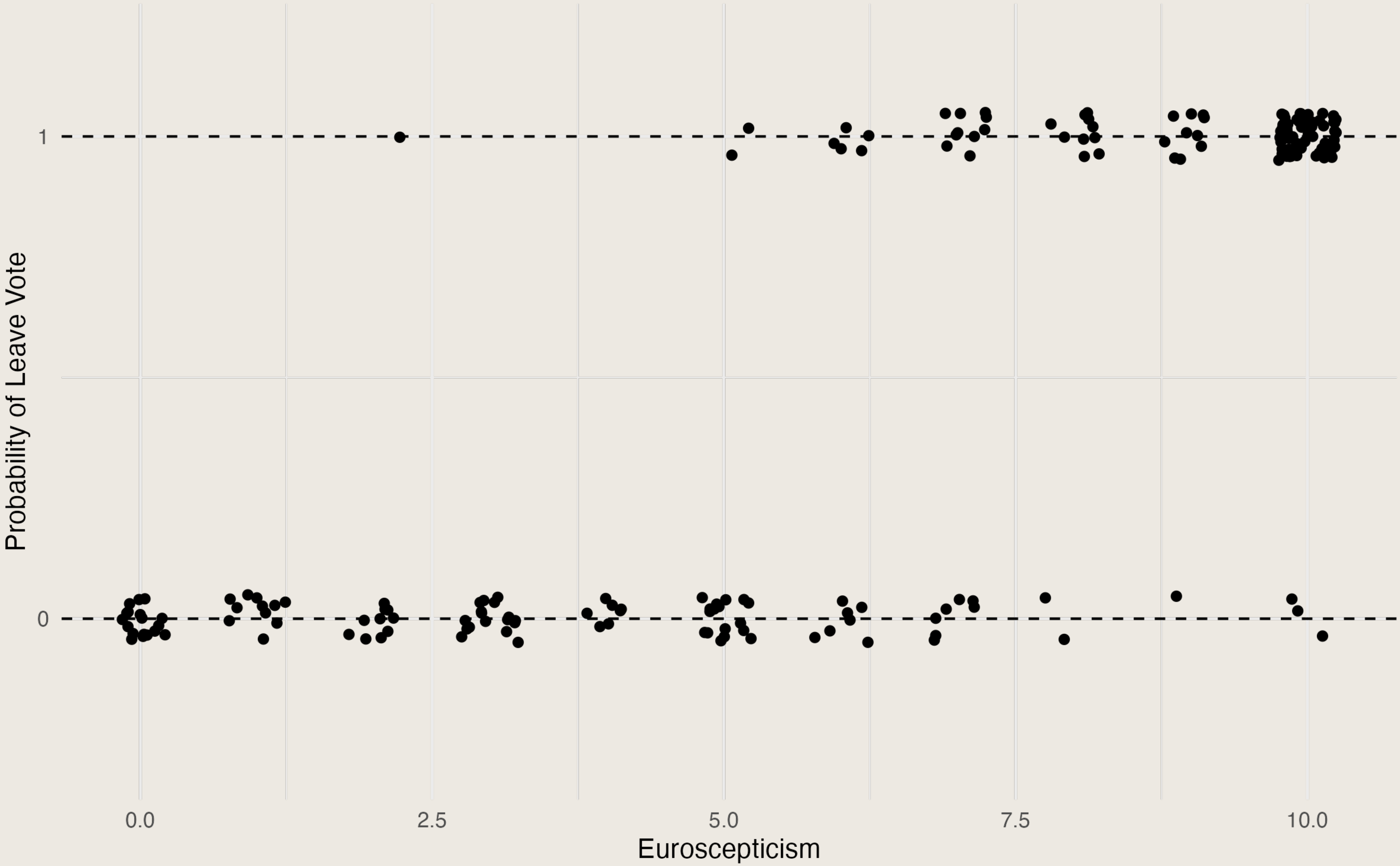
Logistic Regression: Motivation

- What does it mean predicting 'turnout', or 'war' or 'Leave vote'?
- When our dependent variable is binary, we would like to know the **probability** of an outcome (vote, war, Leave vote) from predictors.
- An approach: just use OLS.

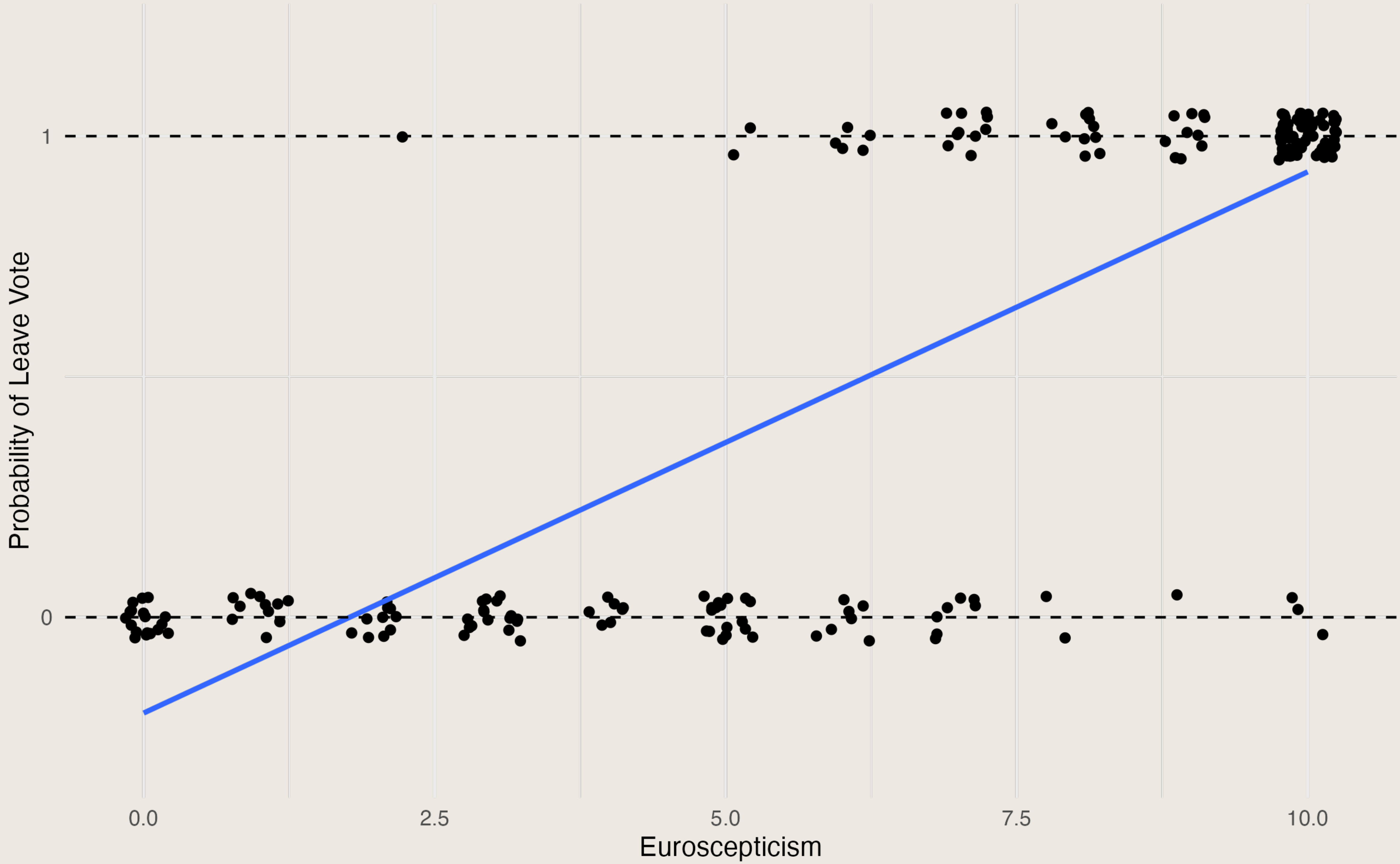
$$\Pr(\text{Leave}_i = 1) = \alpha + \beta \text{Euroscepticism (0-10 scale)}_i + \epsilon_i$$

- **Linear Probability Model (LPM):** regress a 0-1 binary variable on covariates; interpret the predicted values as fractional probabilities.

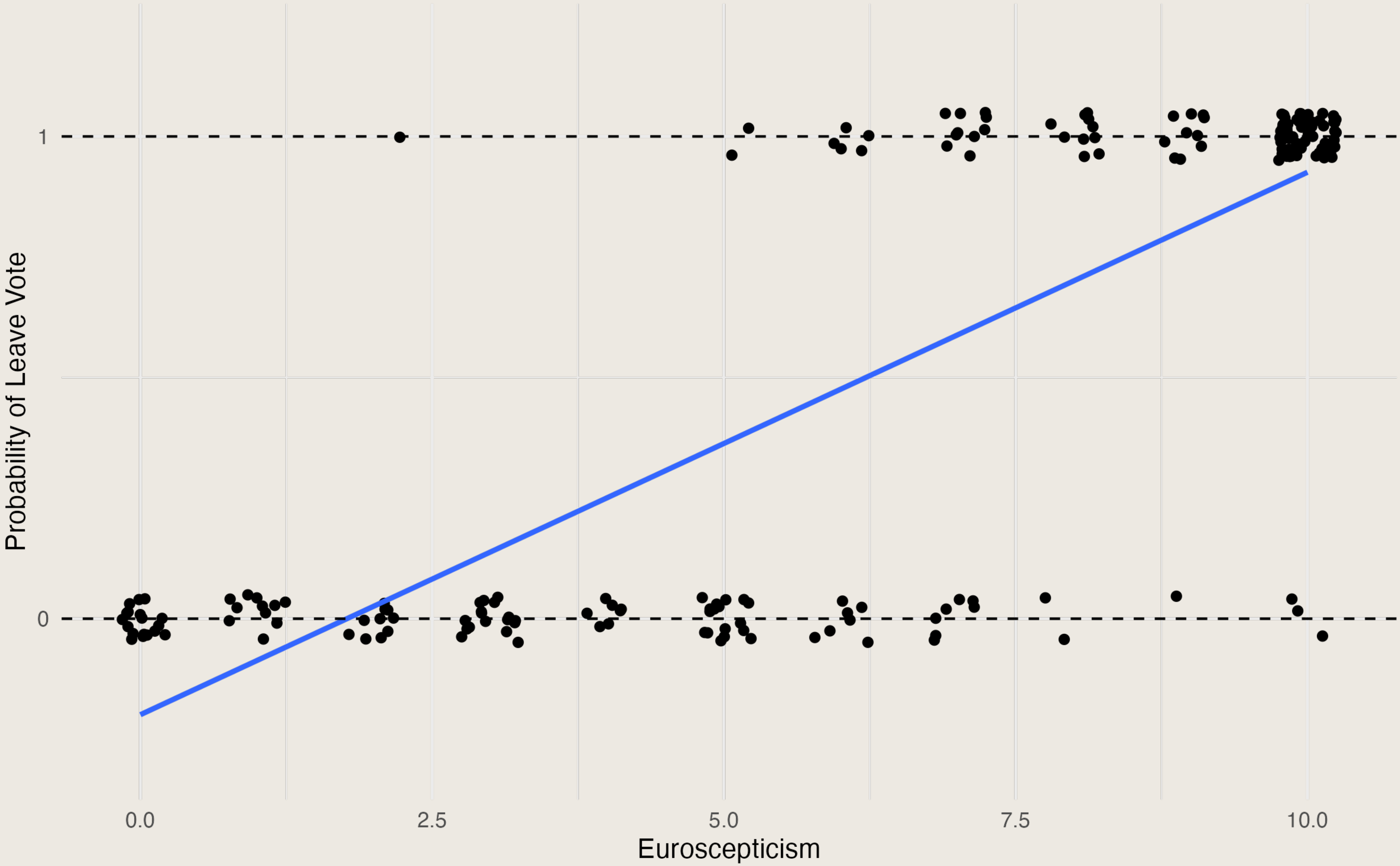
Logistic Regression: Motivation



Logistic Regression: Motivation



Logistic Regression: Motivation



What's wrong with this?

Logistic Regression: Motivation

Logistic Regression: Motivation

- **Unbounded predictions:** LPMs often returns negative probabilities or probabilities above 1.

Logistic Regression: Motivation

* **Unbounded predictions:** LPMs often returns negative probabilities or probabilities above 1.
* More problems:

Logistic Regression: Motivation

- **Unbounded predictions:** LPMs often returns negative probabilities or probabilities above 1.
- More problems:
  - **Non-normal errors:** For any value of $\hat{Y}$, only two possible errors: $0 - \hat{Y}$ if $Y = 0$ and $1 - \hat{Y}$ if $Y = 1$. These won't distribute normally.

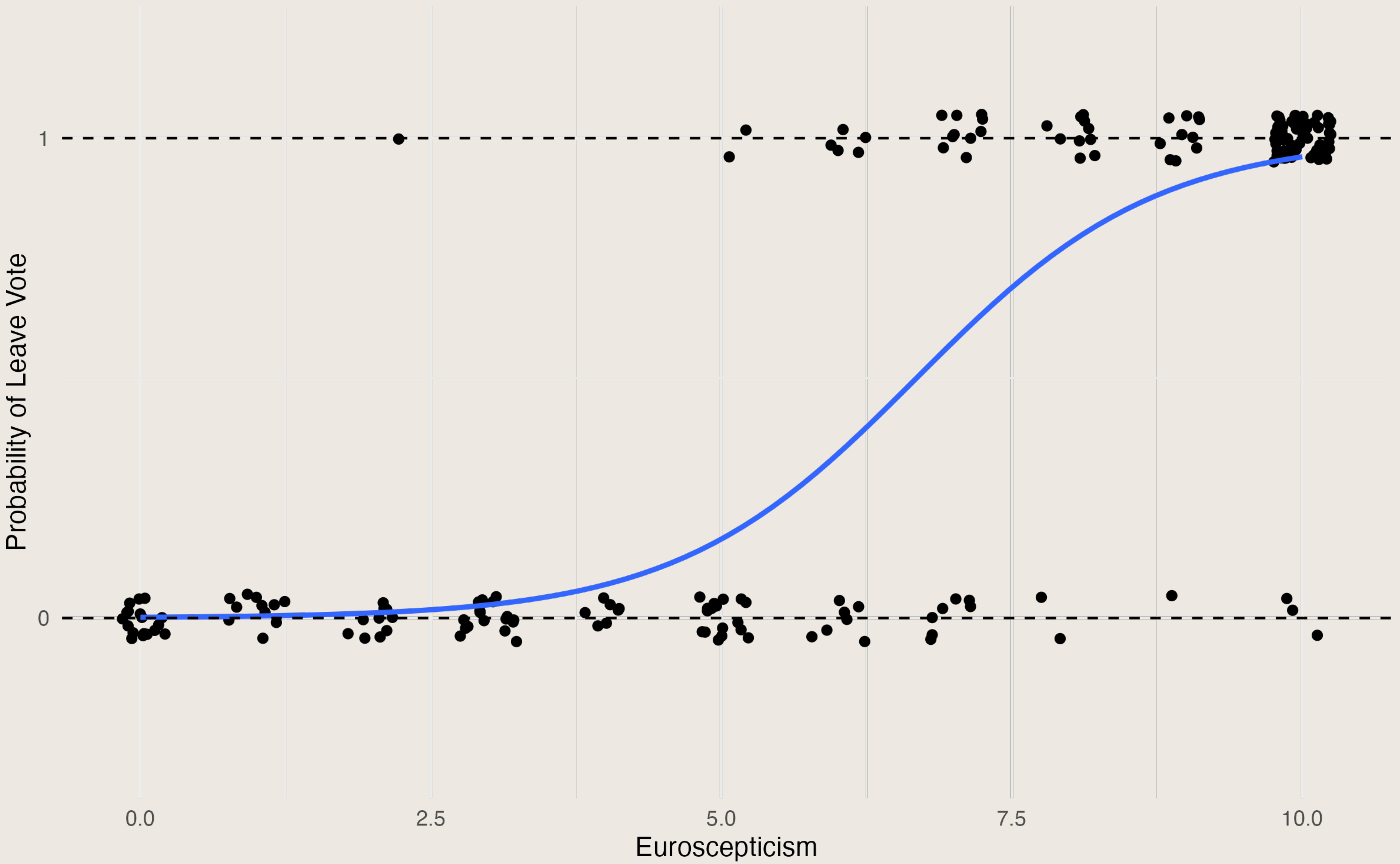
Logistic Regression: Motivation

- **Unbounded predictions:** LPMs often returns negative probabilities or probabilities above 1.
- More problems:
  - **Non-normal errors:** For any value of $\hat{Y}$, only two possible errors: $0 - \hat{Y}$ if $Y = 0$ and $1 - \hat{Y}$ if $Y = 1$. These won't distribute normally.
  - **Non-constant variance:** Y_i is either 0 or 1, but $\hat{Y}_i$ can be any value. So the absolute size of the error $\epsilon_i = Y_i - \hat{Y}_i$ gets smaller as $\hat{Y}$ gets closer to 0 or 1, and bigger as it gets farther. So Var(ϵ) and $\hat{Y}$ are correlated.

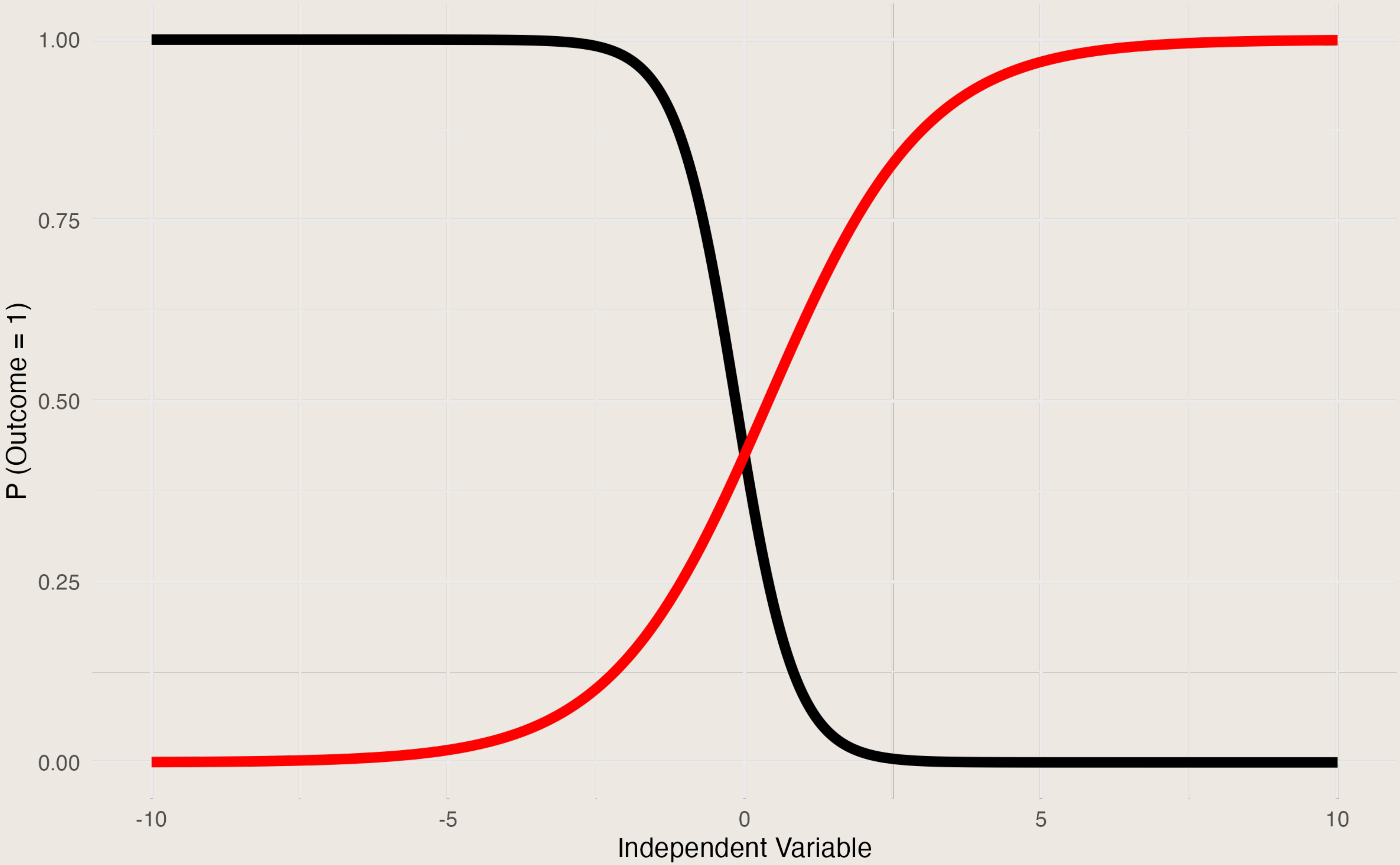
Logistic Regression: Motivation

- **Unbounded predictions:** LPMs often returns negative probabilities or probabilities above 1.
- More problems:
  - **Non-normal errors:** For any value of $\hat{Y}$, only two possible errors: $0 - \hat{Y}$ if $Y = 0$ and $1 - \hat{Y}$ if $Y = 1$. These won't distribute normally.
  - **Non-constant variance:** Y_i is either 0 or 1, but $\hat{Y}_i$ can be any value. So the absolute size of the error $\epsilon_i = Y_i - \hat{Y}_i$ gets smaller as $\hat{Y}$ gets closer to 0 or 1, and bigger as it gets farther. So $\text{Var}(\epsilon)$ and $\hat{Y}$ are correlated.
- Only advantages of LPMs: easy-to-interpret coefficients and computationally faster than alternative. With today's software, generally **no good reason to use them** (though still pop up in econ).

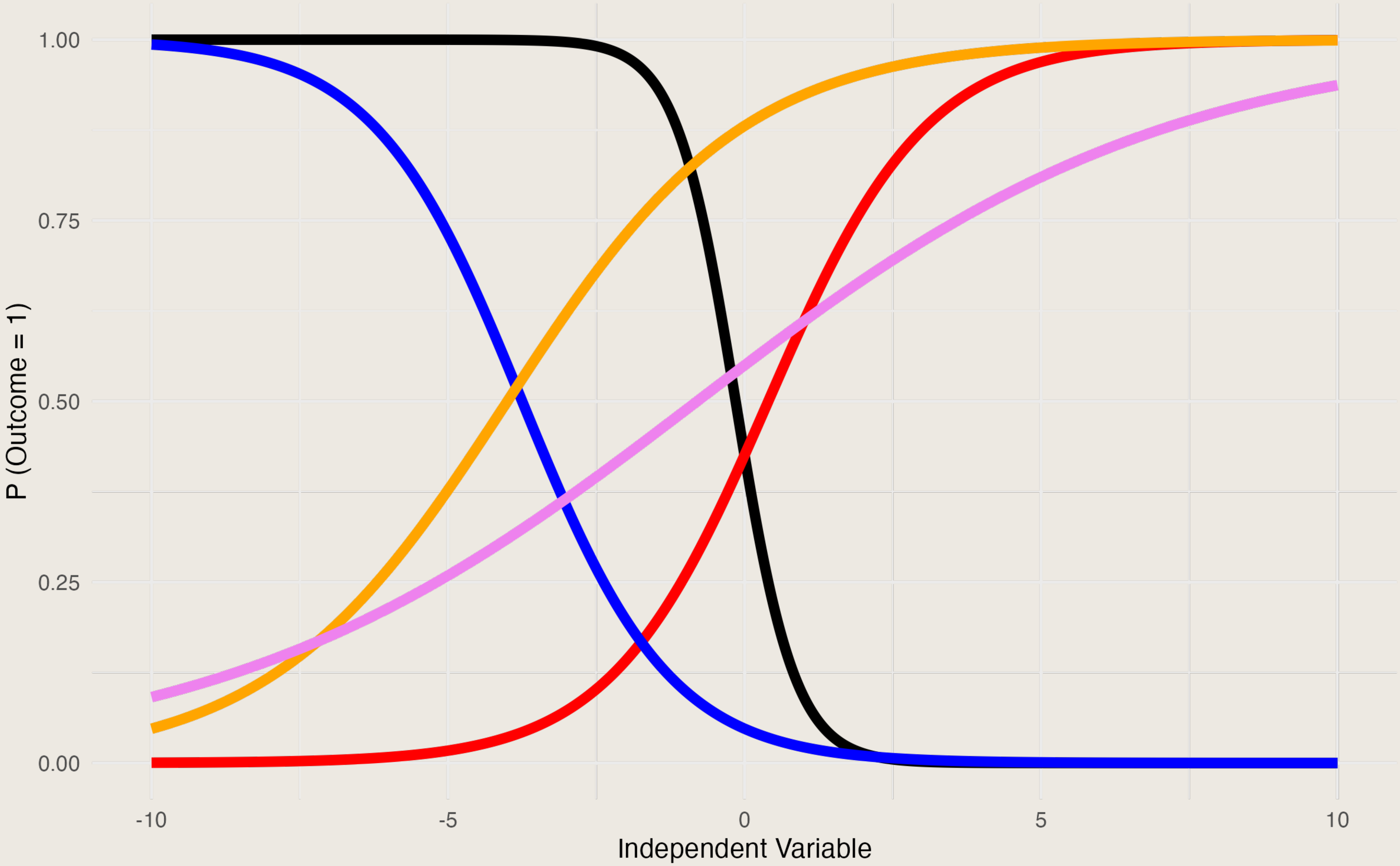
Logistic Regression: Intuition



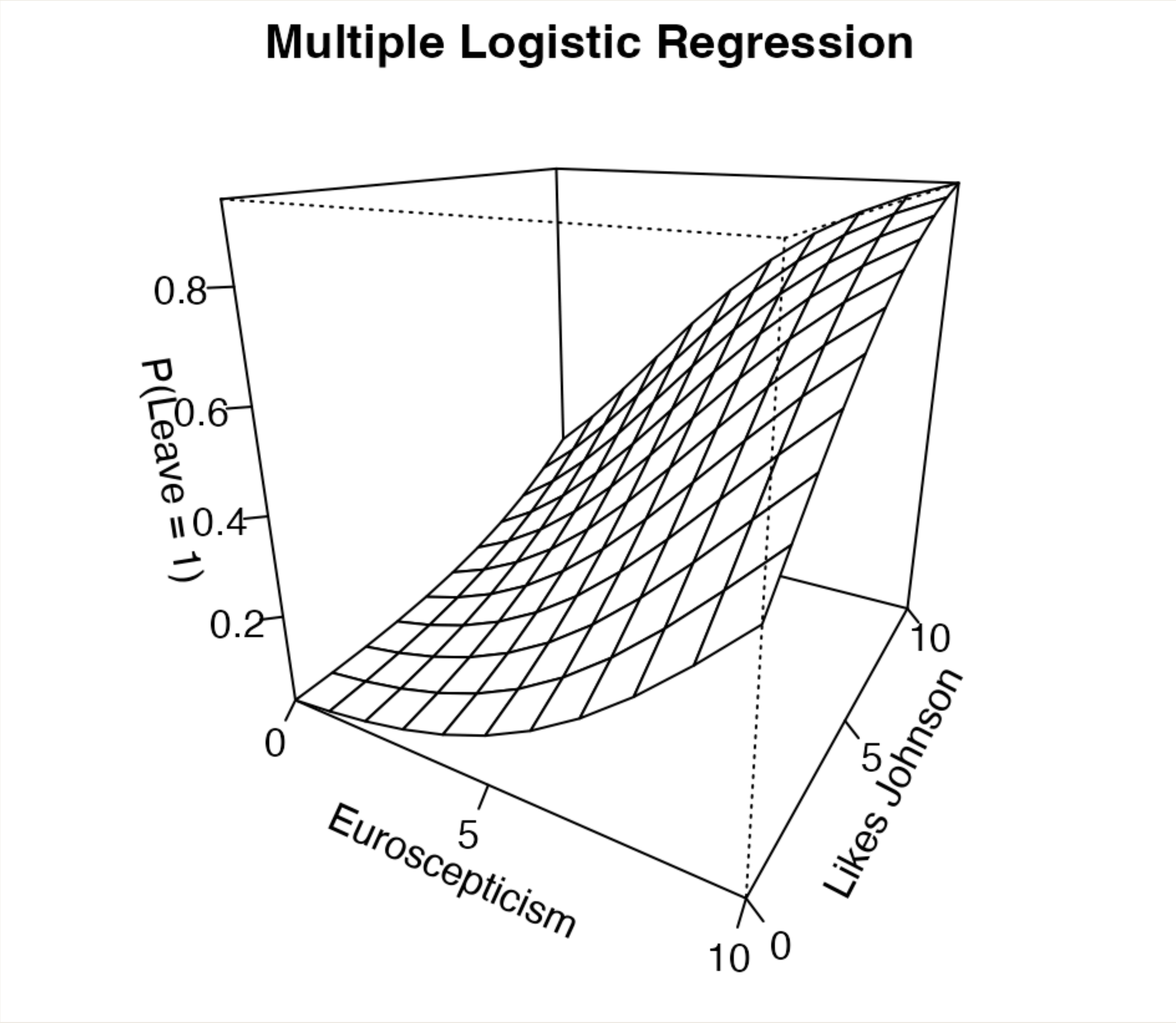
Logistic Regression ‘Squiggles’



Logistic Regression 'Squiggles'



‘Squiggles’ in Multiple Dimensions



Logistic Regression

Logistic Regression

- The 'squiggle' is known as a **sigmoid**. Its equation is:

Logistic Regression

- The 'squiggle' is known as a **sigmoid**. Its equation is:

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

Logistic Regression

- The ‘squiggle’ is known as a **sigmoid**. Its equation is:

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

- Where e is 2.71828… (Euler’s number).

Logistic Regression

- The 'squiggle' is known as a **sigmoid**. Its equation is:

$$\Pr(Y=1) = \frac{1}{1+e^{-(\alpha+\beta X)}}$$

- Where e is 2.71828… (Euler's number).
- Raising e to the power of something is the inverse of taking a natural logarithm of something:

Logistic Regression

- The ‘squiggle’ is known as a **sigmoid**. Its equation is:

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

- Where e is 2.71828… (Euler’s number).
- Raising e to the power of something is the inverse of taking a natural logarithm of something:

$$e^3 = 20.08554... \qquad \rightarrow \qquad \log(20.08554...) = 3$$

Logistic Regression

Logistic Regression

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

Logistic Regression

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

- can be rearranged with a bit of algebra as…

Logistic Regression

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

- can be rearranged with a bit of algebra as…

$$\log \frac{\Pr(Y = 1)}{1 - \Pr(Y = 1)} = \alpha + \beta X$$

Logistic Regression

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

- can be rearranged with a bit of algebra as…

$$\log \frac{\Pr(Y = 1)}{1 - \Pr(Y = 1)} = \alpha + \beta X$$

- And because Y is binary, $1 - \Pr(Y = 1)$ is the same as $\Pr(Y = 0)$…

Logistic Regression

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha+\beta X)}}$$

- can be rearranged with a bit of algebra as…

$$\log \frac{\Pr(Y = 1)}{1 - \Pr(Y = 1)} = \alpha + \beta X$$

- And because Y is binary, $1 - \Pr(Y = 1)$ is the same as $\Pr(Y = 0)$…

$$\log \frac{\Pr(Y = 1)}{\Pr(Y = 0)} = \alpha + \beta X$$

Logistic Regression

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\alpha+\beta X)}}$$

- can be rearranged with a bit of algebra as…

$$\log \frac{\Pr(Y = 1)}{1 - \Pr(Y = 1)} = \alpha + \beta X$$

- And because Y is binary, $1 - \Pr(Y = 1)$ is the same as $\Pr(Y = 0)$…

$$\log \frac{\Pr(Y = 1)}{\Pr(Y = 0)} = \alpha + \beta X$$

- $\log \frac{\Pr(Y = 1)}{1 - \Pr(Y = 1)}$ is known as **log-odds**, or **logit** function of $\Pr(Y = 1)$.

Log-Odds

Log-Odds

Odds are a way to express probability (used in betting). Odds $= \dfrac{p}{1-p}$

Log-Odds

Odds are a way to express probability (used in betting). Odds $= \dfrac{p}{1-p}$

* For $p = 0.5$ (50% probability), odds is 1.

Log-Odds

Odds are a way to express probability (used in betting). Odds $= \dfrac{p}{1-p}$

- For $p = 0.5$ (50% probability), odds is 1.
- For $p = 0.75$, odds is 3, because $0.75/(1-0.75) = 3$

Log-Odds

Odds are a way to express probability (used in betting). Odds = $\dfrac{p}{1-p}$

- For $p = 0.5$ (50% probability), odds is 1.
- For $p = 0.75$, odds is 3, because $0.75/(1-0.75) = 3$
- For $p = 0.25$, odds is $1/3$, because $0.25/(1-0.25) = 1/3$

Log-Odds

Odds are a way to express probability (used in betting). Odds = $\dfrac{p}{1-p}$

- For $p = 0.5$ (50% probability), odds is 1.
- For $p = 0.75$, odds is 3, because $0.75/(1-0.75) = 3$
- For $p = 0.25$, odds is $1/3$, because $0.25/(1-0.25) = 1/3$

Log-odds centre the outcome at 0 and linearise it:

Log-Odds

Odds are a way to express probability (used in betting). Odds $= \dfrac{p}{1-p}$

- For $p = 0.5$ (50% probability), odds is 1.
- For $p = 0.75$, odds is 3, because $0.75/(1 - 0.75) = 3$
- For $p = 0.25$, odds is $1/3$, because $0.25/(1 - 0.25) = 1/3$

Log-odds centre the outcome at 0 and linearise it:

- For $p = 0.5$, **odds** is $1 \rightarrow$ **log-odds or logit**(0.5) is $\log(1) = 0$

Log-Odds

Odds are a way to express probability (used in betting). Odds = $\frac{p}{1-p}$

- For $p = 0.5$ (50% probability), odds is 1.
- For $p = 0.75$, odds is 3, because $0.75/(1-0.75) = 3$
- For $p = 0.25$, odds is $1/3$, because $0.25/(1-0.25) = 1/3$

Log-odds centre the outcome at 0 and linearise it:

- For $p = 0.5$, **odds** is $1 \rightarrow$ **log-odds or logit**(0.5) is $\log(1) = 0$
- For $p = 0.75$, **odds** is $3 \rightarrow$ **log-odds or logit**(0.75) is $\log(3) \approx 1.10$

Log-Odds

Odds are a way to express probability (used in betting). Odds = $\dfrac{p}{1-p}$

* For $p = 0.5$ (50% probability), odds is 1.
* For $p = 0.75$, odds is 3, because $0.75/(1-0.75) = 3$
* For $p = 0.25$, odds is $1/3$, because $0.25/(1-0.25) = 1/3$

Log-odds centre the outcome at 0 and linearise it:

* For $p = 0.5$, **odds** is $1 \rightarrow$ **log-odds or logit**(0.5) is $\log(1) = 0$
* For $p = 0.75$, **odds** is $3 \rightarrow$ **log-odds or logit**(0.75) is $\log(3) \approx 1.10$
* For $p = 0.25$, **odds** is $1/3 \rightarrow$ **log-odds or logit**(0.25) is $\log(1/3) \approx -1.10$

Log-Odds

Odds are a way to express probability (used in betting). Odds $= \dfrac{p}{1-p}$

* For $p = 0.5$ (50% probability), odds is 1.
* For $p = 0.75$, odds is 3, because $0.75/(1 - 0.75) = 3$
* For $p = 0.25$, odds is $1/3$, because $0.25/(1 - 0.25) = 1/3$

Log-odds centre the outcome at 0 and linearise it:

* For $p = 0.5$, **odds** is $1 \rightarrow$ **log-odds or logit**(0.5) is $\log(1) = 0$
* For $p = 0.75$, **odds** is $3 \rightarrow$ **log-odds or logit**(0.75) is $\log(3) \approx 1.10$
* For $p = 0.25$, **odds** is $1/3 \rightarrow$ **log-odds or logit**(0.25) is $\log(1/3) \approx -1.10$
* For $p = 0.9$, **odds** is $9 \rightarrow$ **log-odds or logit**(0.9) is $\log(9) \approx 2.20$

Log-Odds

Odds are a way to express probability (used in betting). Odds = $\dfrac{p}{1-p}$

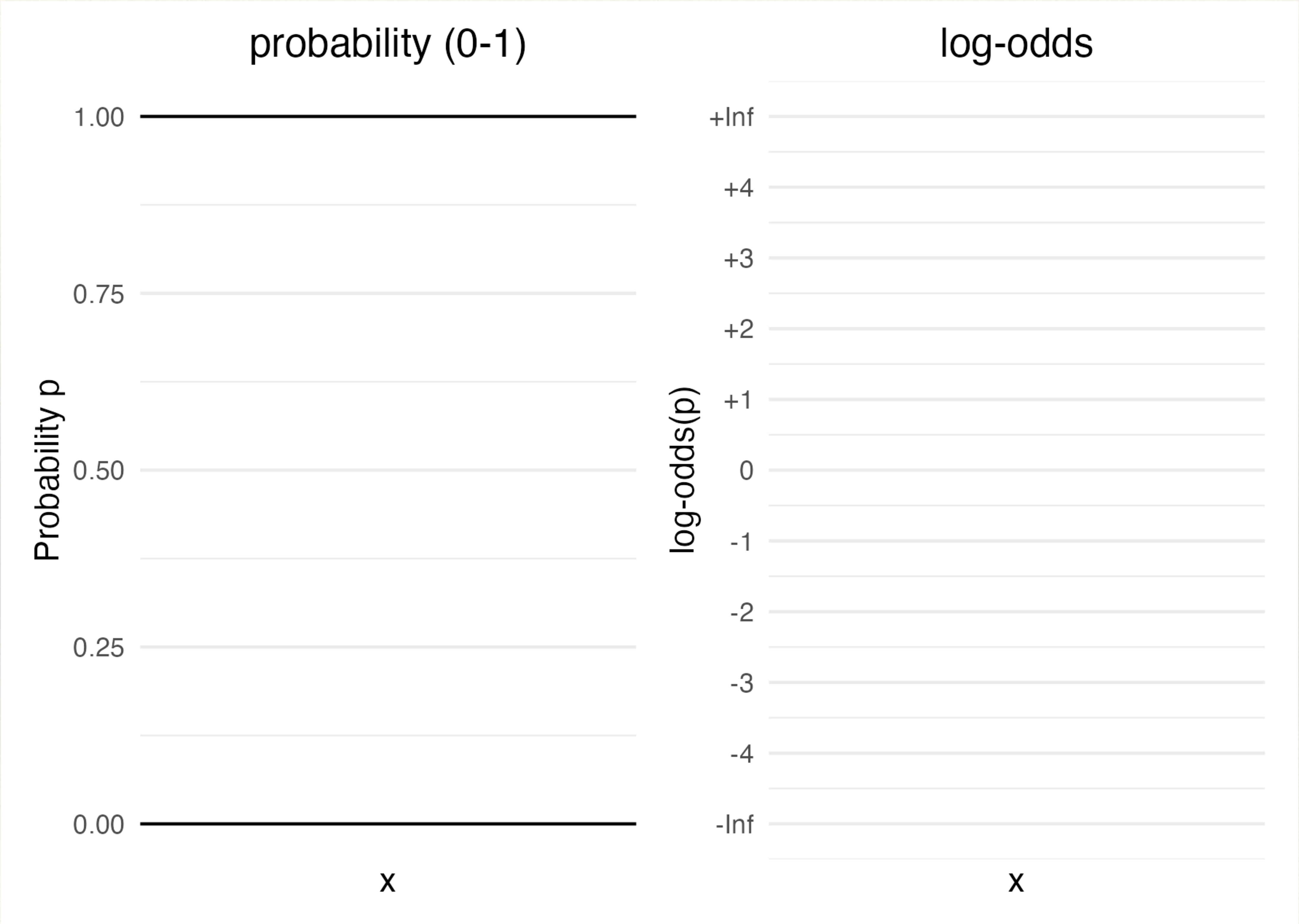
- For $p = 0.5$ (50% probability), odds is 1.
- For $p = 0.75$, odds is 3, because $0.75/(1-0.75) = 3$
- For $p = 0.25$, odds is $1/3$, because $0.25/(1-0.25) = 1/3$

Log-odds centre the outcome at 0 and linearise it:

- For $p = 0.5$, **odds** is $1 \rightarrow$ **log-odds or logit**(0.5) is $\log(1) = 0$
- For $p = 0.75$, **odds** is $3 \rightarrow$ **log-odds or logit**(0.75) is $\log(3) \approx 1.10$
- For $p = 0.25$, **odds** is $1/3 \rightarrow$ **log-odds or logit**(0.25) is $\log(1/3) \approx -1.10$
- For $p = 0.9$, **odds** is $9 \rightarrow$ **log-odds or logit**(0.9) is $\log(9) \approx 2.20$
- For $p = 0.1$, **odds** is $1/9 \rightarrow$ **log-odds or logit**(0.1) is $\log(1/9) \approx -2.20$

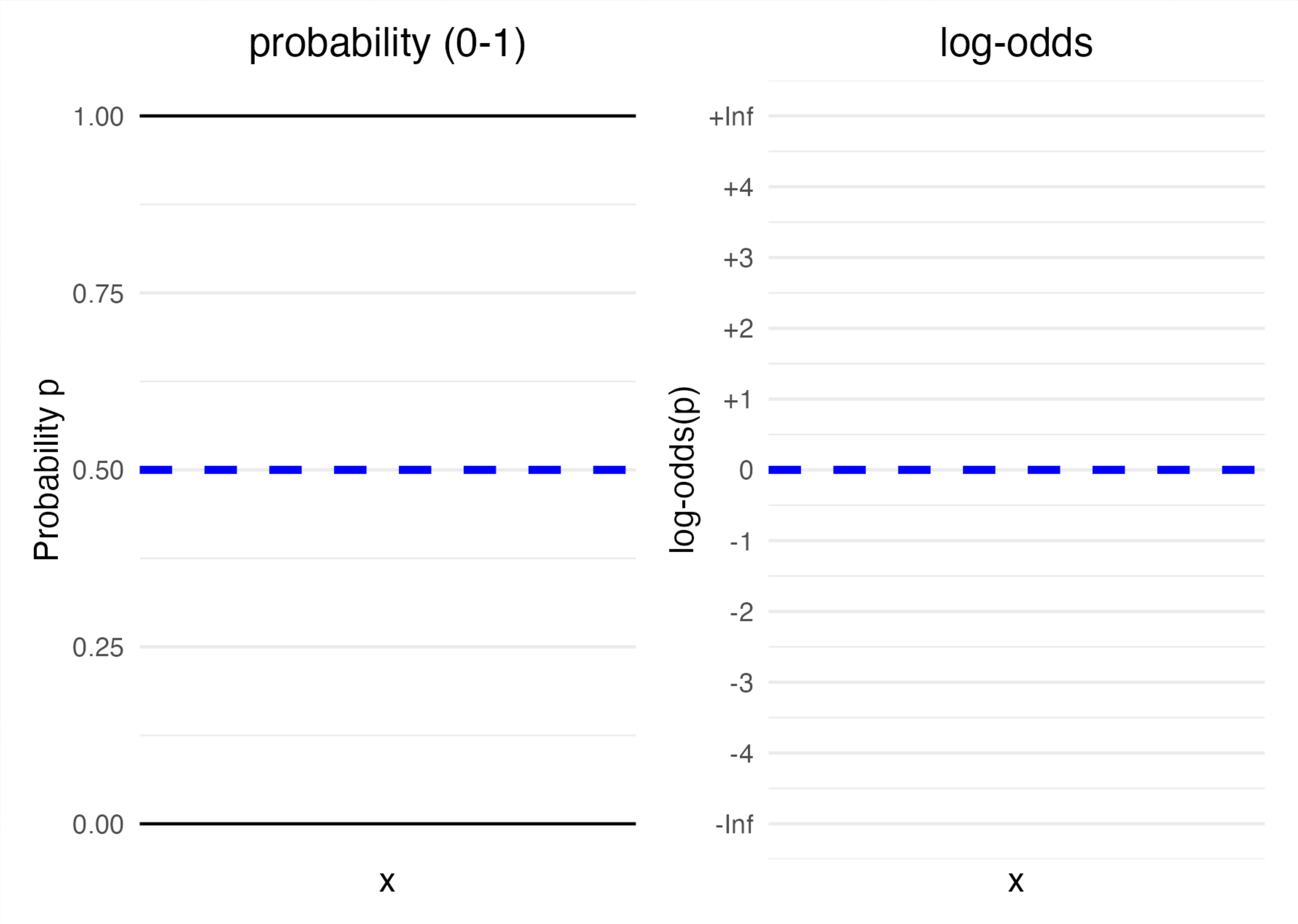
$$p = \frac{1}{1 + e^{-x}}$$

$$\log\left(\frac{p}{1-p}\right) = x$$



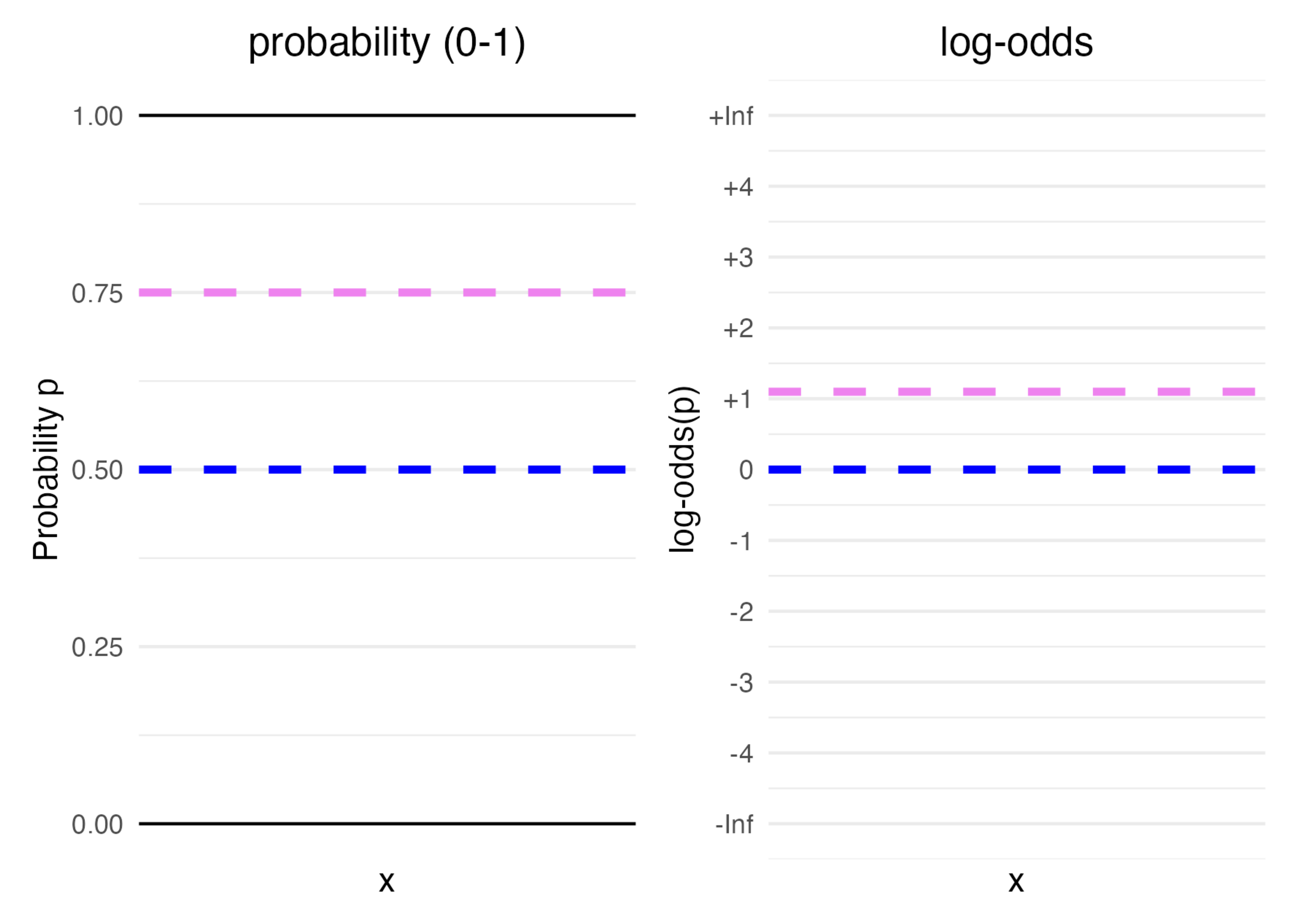
$$p = \frac{1}{1 + e^{-x}}$$

$$\log\left(\frac{p}{1-p}\right) = x$$



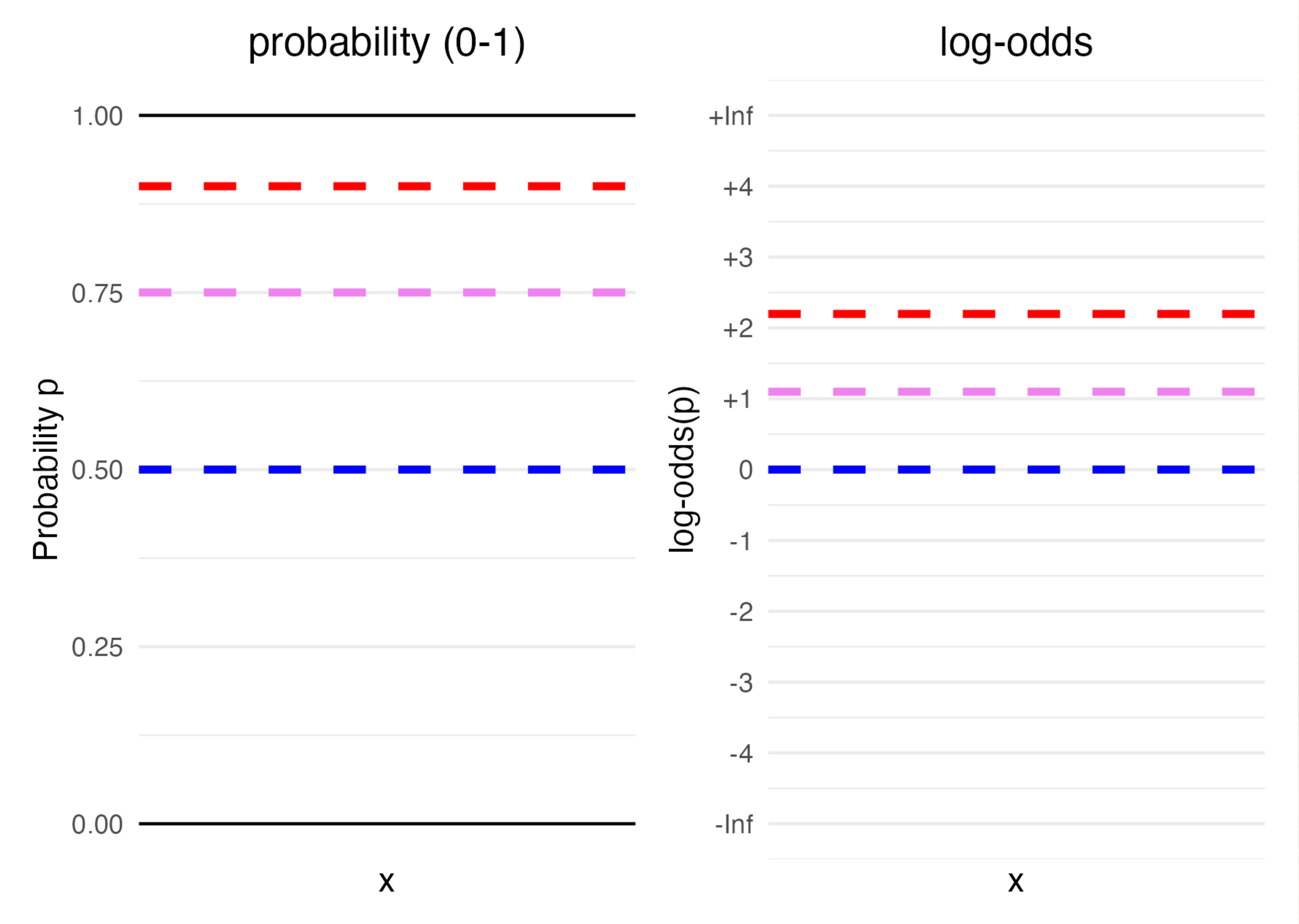
$$p = \frac{1}{1 + e^{-x}}$$

$$\log\left(\frac{p}{1-p}\right) = x$$

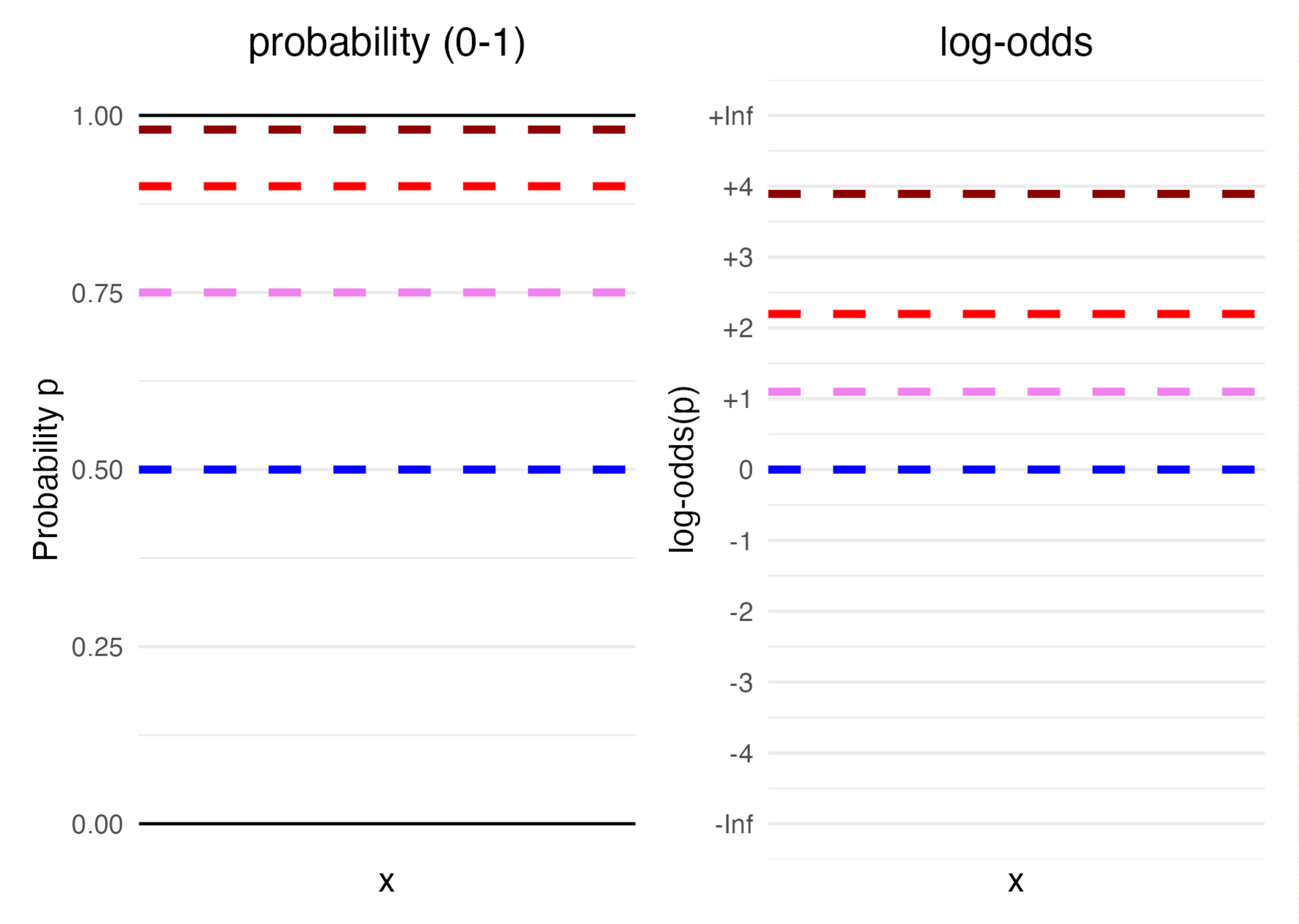


$$p = \frac{1}{1 + e^{-x}}$$

$$\log\left(\frac{p}{1-p}\right) = x$$

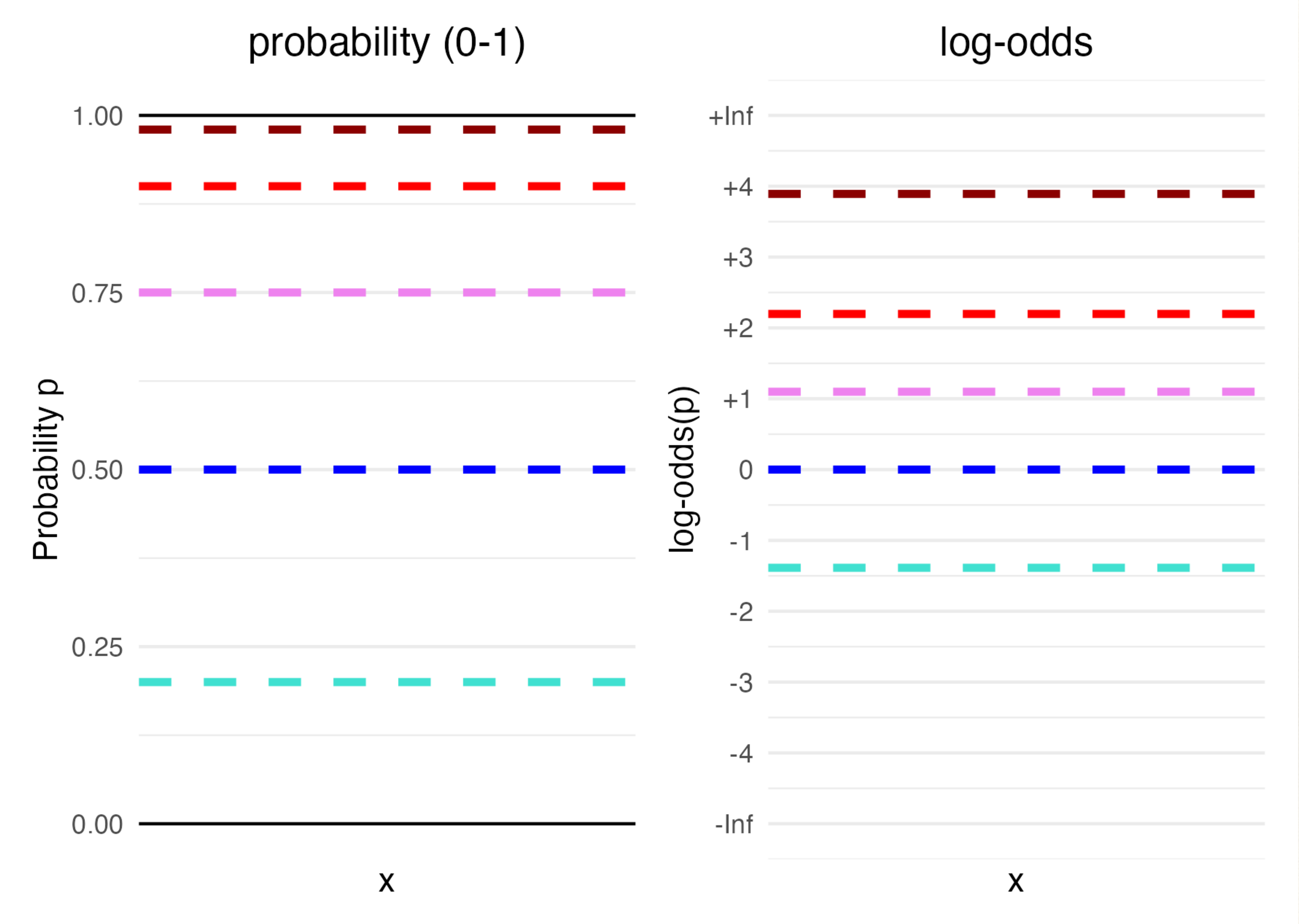


$$p = \frac{1}{1+e^{-x}} \qquad \log\left(\frac{p}{1-p}\right) = x$$



$$p = \frac{1}{1 + e^{-x}}$$

$$\log\left(\frac{p}{1-p}\right) = x$$



Probability to Log-Odds and Back

Probability to Log-Odds and Back

- If I have a probability p, I can get the log-odds with the **logit function**:

Probability to Log-Odds and Back

* If I have a probability *p*, I can get the log-odds with the **logit function**:

$$\text{log-odds} = \log(\frac{p}{1-p}) = \text{logit}(p)$$

Probability to Log-Odds and Back

- If I have a probability *p*, I can get the log-odds with the **logit function**:

$$\text{log-odds} = \log\left(\frac{p}{1-p}\right) = \text{logit}(p)$$

- If I have the log-odds *x*, and I want the probability, I use the **logistic function** (also known as **inverse-logit**, and notated as logit^{-1}):

Probability to Log-Odds and Back

* If I have a probability p, I can get the log-odds with the **logit function**:

$$\text{log-odds} = \log(\frac{p}{1-p}) = \text{logit}(p)$$

* If I have the log-odds x, and I want the probability, I use the **logistic function** (also known as **inverse-logit**, and notated as logit^{-1}):

$$p = \frac{1}{1+e^{-x}} = \text{logit}^{-1}(x)$$

Probability to Log-Odds and Back

- If I have a probability p, I can get the log-odds with the **logit function**:

$$\text{log-odds} = \log\left(\frac{p}{1-p}\right) = \text{logit}(p)$$

- If I have the log-odds x, and I want the probability, I use the **logistic function** (also known as **inverse-logit**, and notated as logit^{-1}):

$$p = \frac{1}{1+e^{-x}} = \text{logit}^{-1}(x)$$

- $\text{logit}(0.25) \approx -1.10$
- $\text{logit}^{-1}(-1.10) \approx 0.25$

Probability to Log-Odds and Back

- If I have a probability p, I can get the log-odds with the **logit function**:

$$\text{log-odds} = \log(\frac{p}{1-p}) = \text{logit}(p)$$

- If I have the log-odds x, and I want the probability, I use the **logistic function** (also known as **inverse-logit**, and notated as logit^{-1}):

$$p = \frac{1}{1+e^{-x}} = \text{logit}^{-1}(x)$$

- $\text{logit}(0.25) \approx -1.10$
- $\text{logit}(0.5) = 0$
- $\text{logit}^{-1}(-1.10) \approx 0.25$
- $\text{logit}^{-1}(0) = 0.5$

Probability to Log-Odds and Back

- If I have a probability p, I can get the log-odds with the **logit function**:

$$\text{log-odds} = \log(\frac{p}{1-p}) = \text{logit}(p)$$

- If I have the log-odds x, and I want the probability, I use the **logistic function** (also known as **inverse-logit**, and notated as logit^{-1}):

$$p = \frac{1}{1+e^{-x}} = \text{logit}^{-1}(x)$$

- $\text{logit}(0.25) \approx -1.10$
- $\text{logit}(0.5) = 0$
- $\text{logit}(0.75) \approx 1.10$
- $\text{logit}^{-1}(-1.10) \approx 0.25$
- $\text{logit}^{-1}(0) = 0.5$
- $\text{logit}^{-1}(1.10) \approx 0.75$

Probability to Log-Odds and Back

- If I have a probability p, I can get the log-odds with the **logit function**:

$$\text{log-odds} = \log\left(\frac{p}{1-p}\right) = \text{logit}(p)$$

- If I have the log-odds x, and I want the probability, I use the **logistic function** (also known as **inverse-logit**, and notated as logit^{-1}):

$$p = \frac{1}{1+e^{-x}} = \text{logit}^{-1}(x)$$

- $\text{logit}(0.25) \approx -1.10$
- $\text{logit}(0.5) = 0$
- $\text{logit}(0.75) \approx 1.10$
- $\text{logit}(0.952) \approx 3$
- $\text{logit}^{-1}(-1.10) \approx 0.25$
- $\text{logit}^{-1}(0) = 0.5$
- $\text{logit}^{-1}(1.10) \approx 0.75$
- $\text{logit}^{-1}(3) \approx 0.952$

Logistic Regression, Two Ways

Logistic Regression, Two Ways

With Y as a probability:

$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha + \beta \text{Euroscepticism})}}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta\text{Euroscepticism})$$

- Easy-to-interpret left-hand side: it's a probability, can only take values comprised between 0 and 1.
- Hard-to-interpret right-hand side: it's a non-linear curve (sigmoid). Not obvious what a 1-unit increase in X does.

Logistic Regression, Two Ways

With Y as a probability:

$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha + \beta \text{Euroscepticism})}}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta \text{Euroscepticism})$$

- Easy-to-interpret left-hand side: it's a probability, can only take values comprised between 0 and 1.
- Hard-to-interpret right-hand side: it's a non-linear curve (sigmoid). Not obvious what a 1-unit increase in X does.

With Y as log-odds:

$$\log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta \text{Euroscepticism}$$

$$\text{logit}[\Pr(\text{Leave} = 1)] = \alpha + \beta \text{Euroscepticism}$$

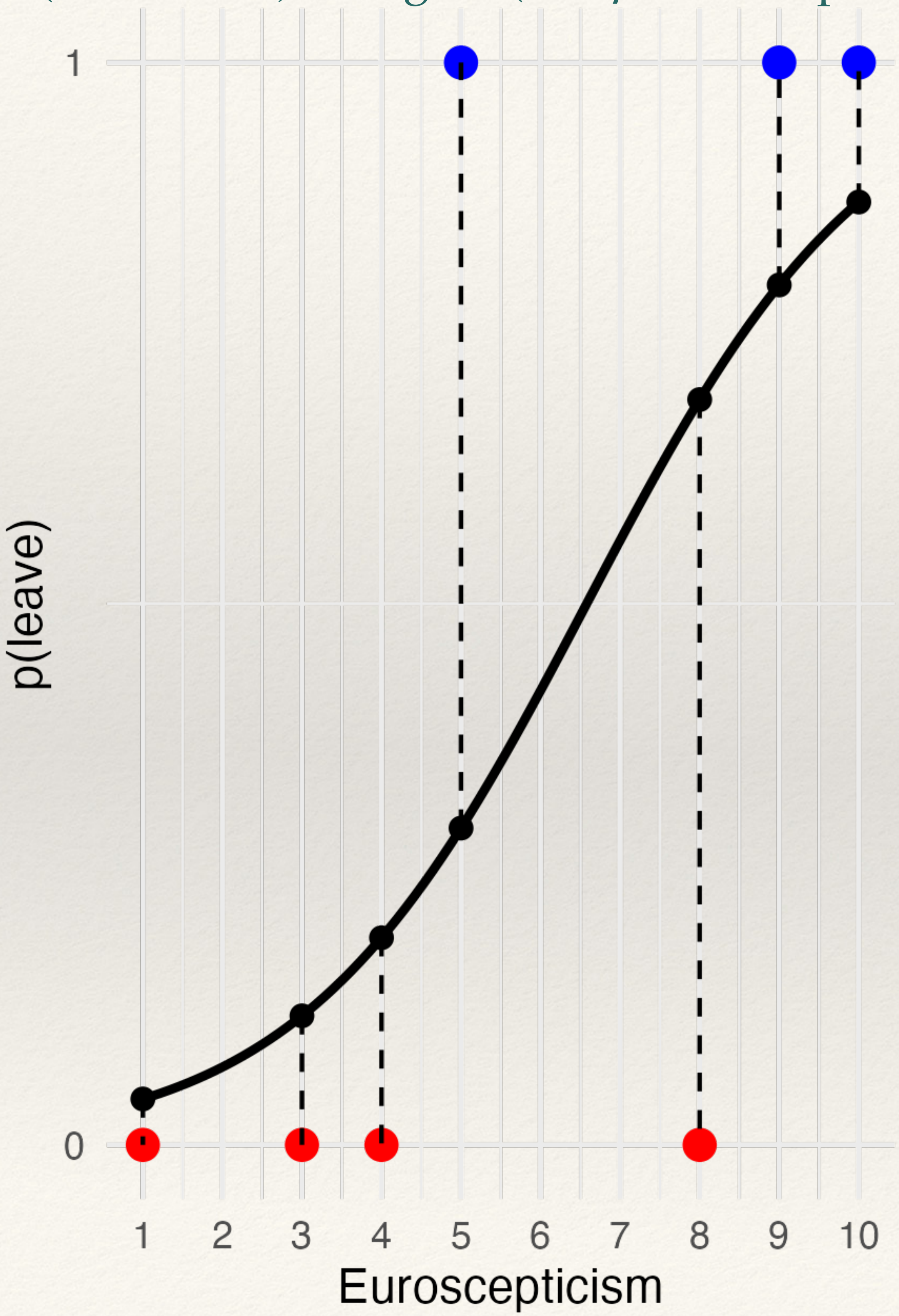
- Easy-to-interpret right-hand side: it's a linear function, like with the linear model. A 1-unit increase in X increases outcome by β.
- Hard-to-interpret left-hand side: it's a funky way of expressing probabilities, which can take any value from $-\inf$ to $+\inf$.

$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha + \beta \text{Euroscepticism})}}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta \text{Euroscepticism})$$

$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha + \beta \text{Euroscepticism})}}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta \text{Euroscepticism})$$

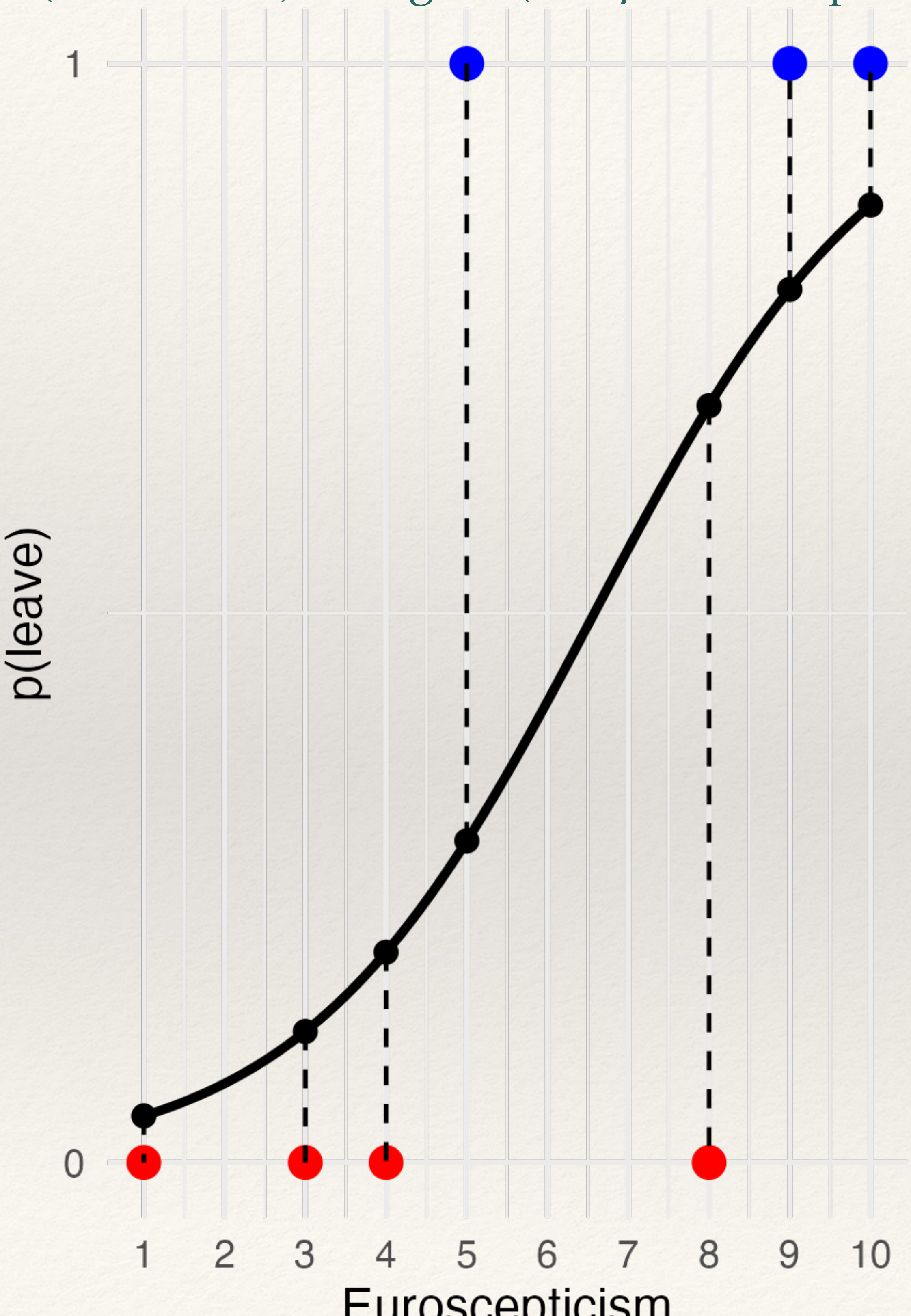


$$\Pr(\text{Leave}=1)=\frac{1}{1+e^{-(\alpha+\beta\text{Euroscepticism})}}$$

$$\log\frac{\Pr(\text{Leave}=1)}{\Pr(\text{Leave}=0)}=\alpha+\beta\text{Euroscepticism}$$

$$\Pr(\text{Leave}=1)=\text{logit}^{-1}(\alpha+\beta\text{Euroscepticism})$$

$$\text{logit}[\Pr(\text{Leave}=1)]=\alpha+\beta\text{Euroscepticism}$$

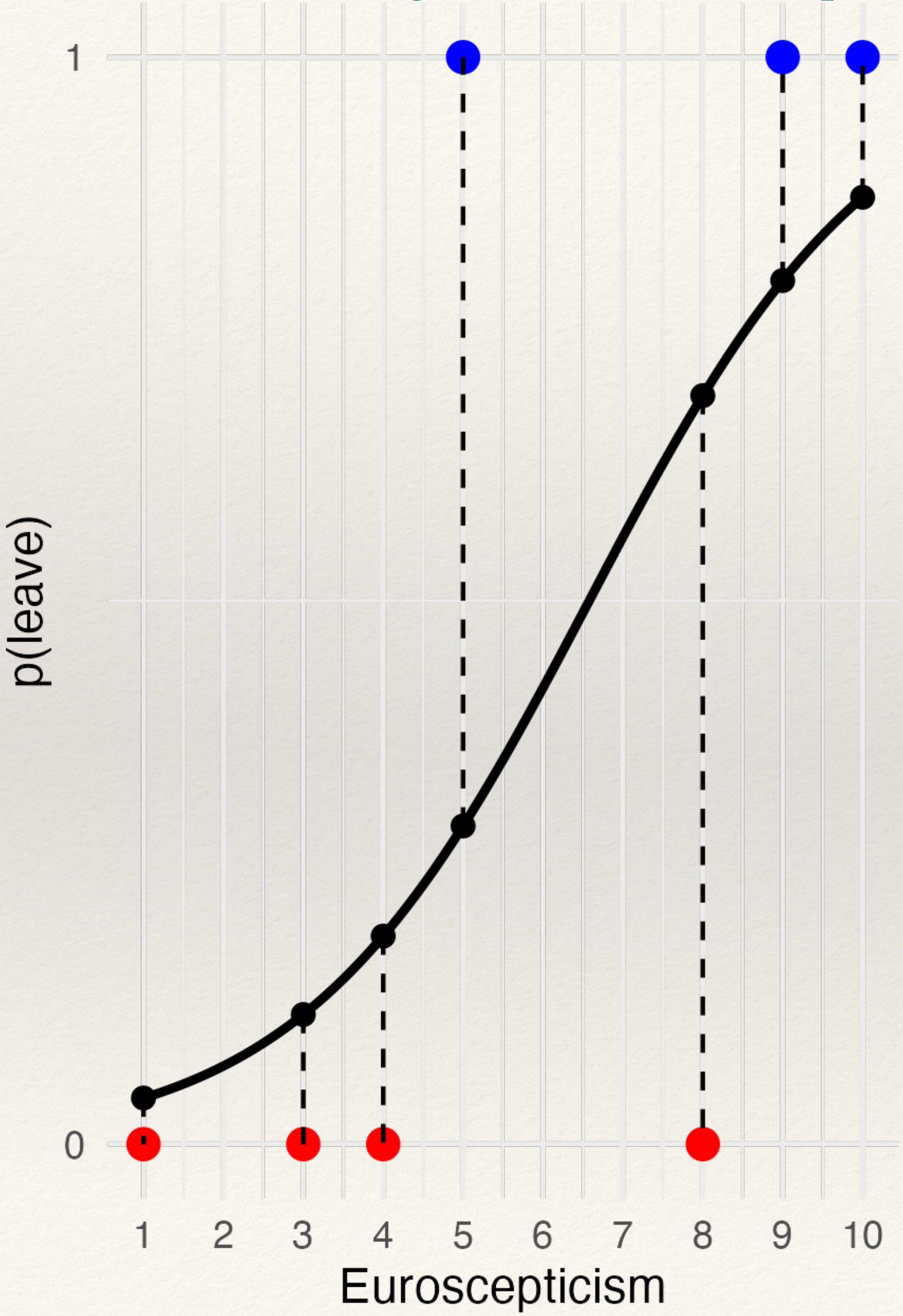


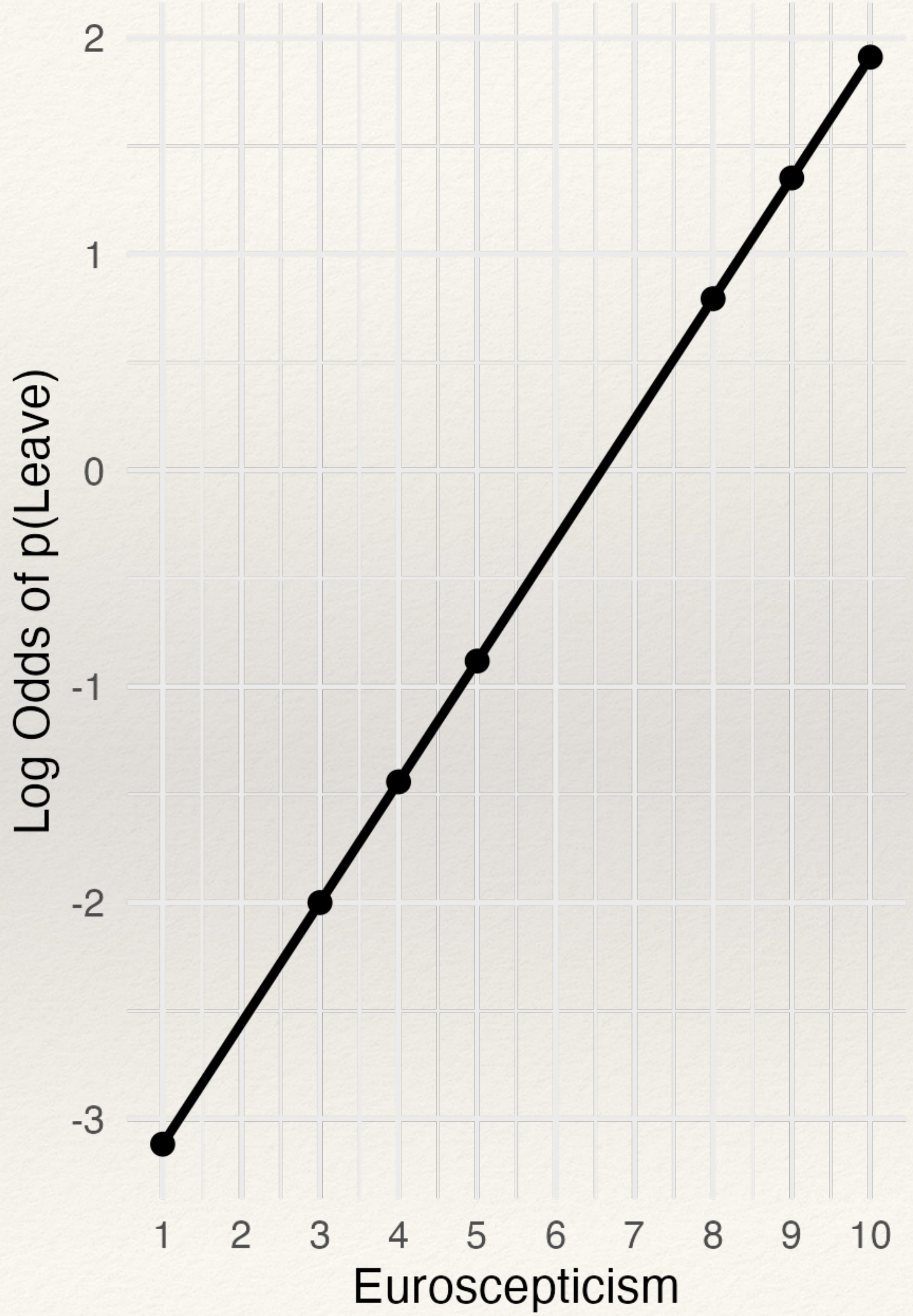
$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha + \beta \text{Euroscepticism})}}$$

$$\log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta \text{Euroscepticism}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta \text{Euroscepticism})$$

$$\text{logit}[\Pr(\text{Leave} = 1)] = \alpha + \beta \text{Euroscepticism}$$



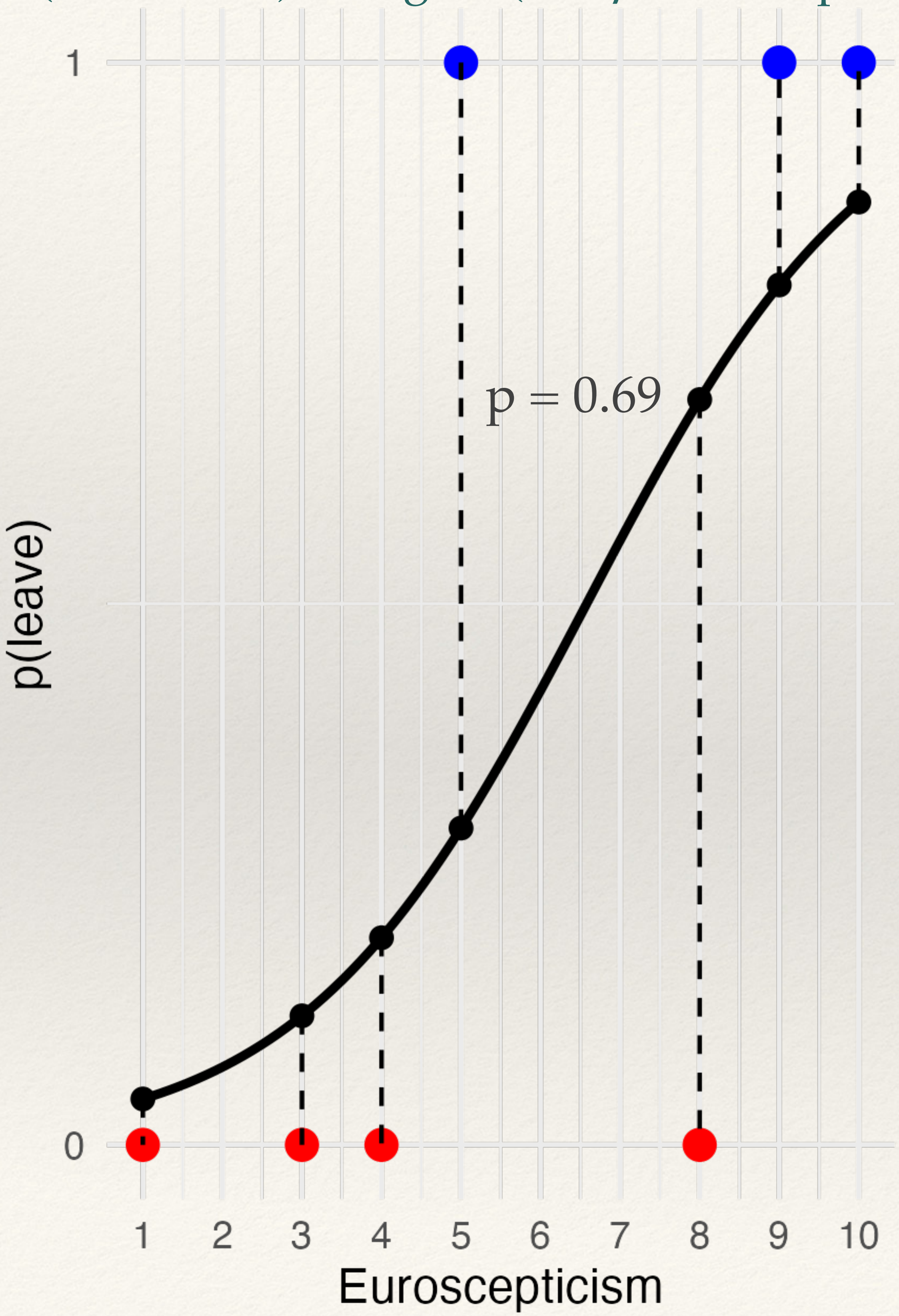


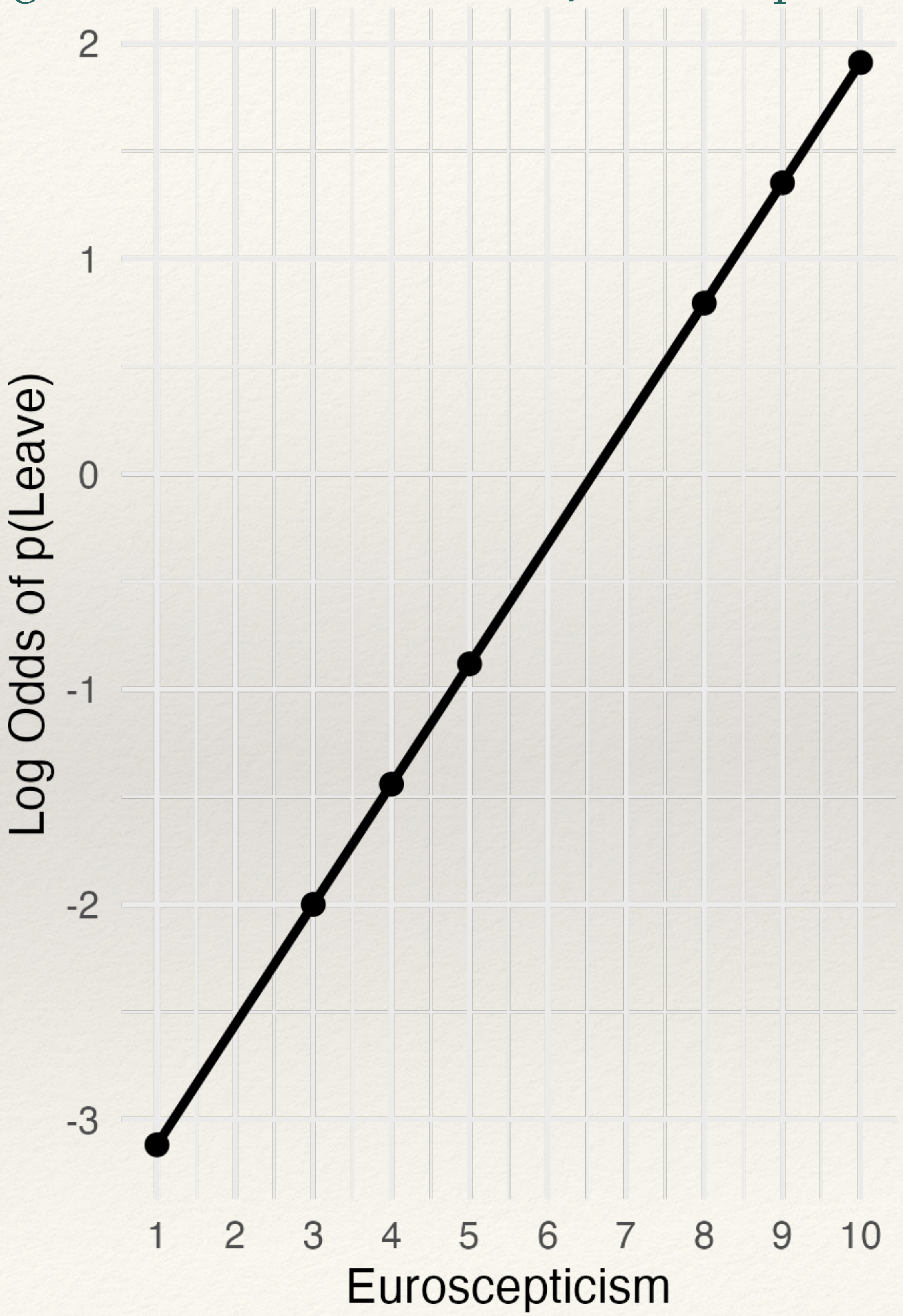
$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha+\beta\text{Euroscepticism})}}$$

$$\log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta\text{Euroscepticism}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta\text{Euroscepticism})$$

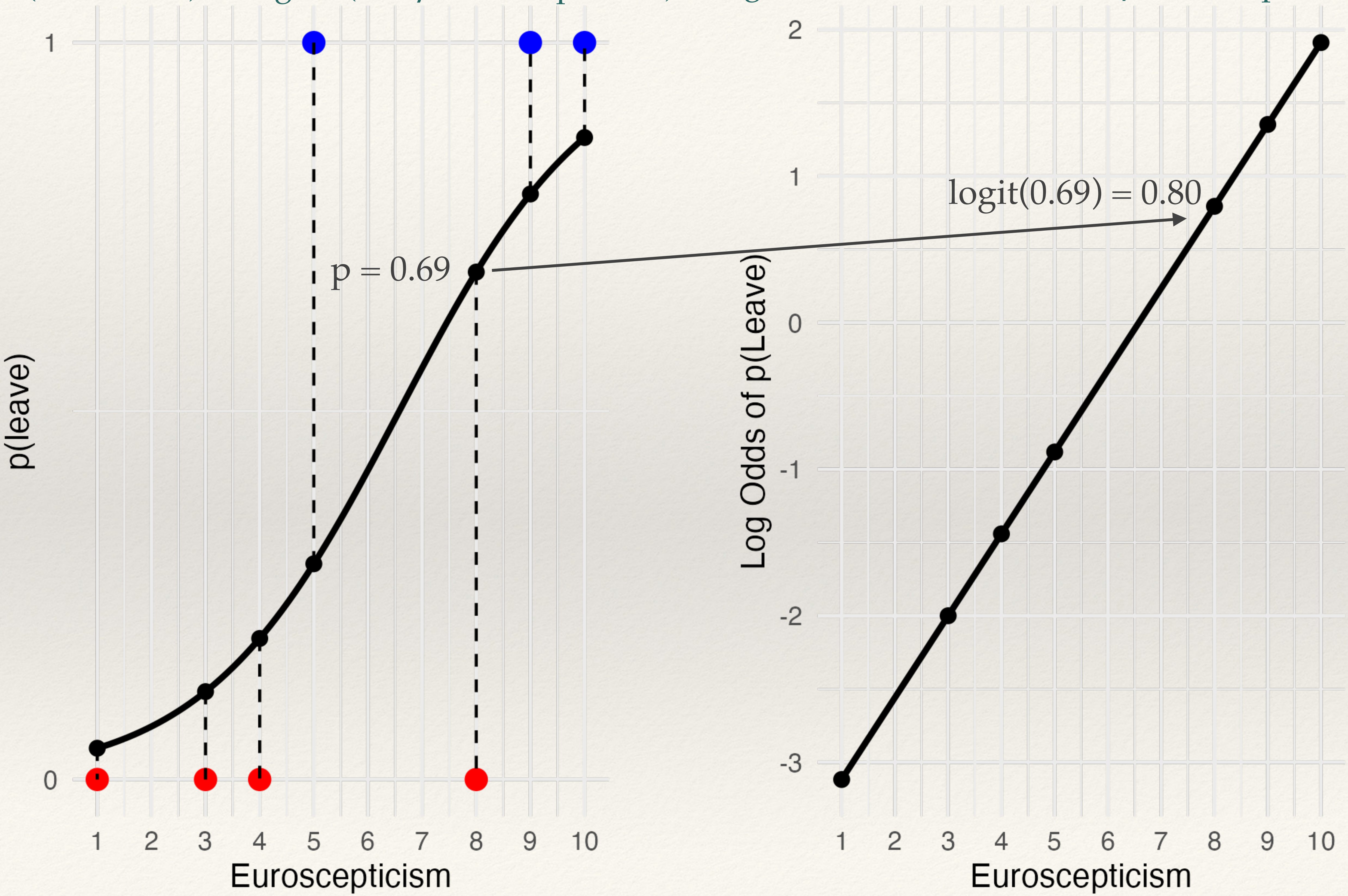
$$\text{logit}[\Pr(\text{Leave} = 1)] = \alpha + \beta\text{Euroscepticism}$$





$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha + \beta \text{Euroscepticism})}} \qquad \log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta \text{Euroscepticism}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta \text{Euroscepticism}) \qquad \text{logit}[\Pr(\text{Leave} = 1)] = \alpha + \beta \text{Euroscepticism}$$

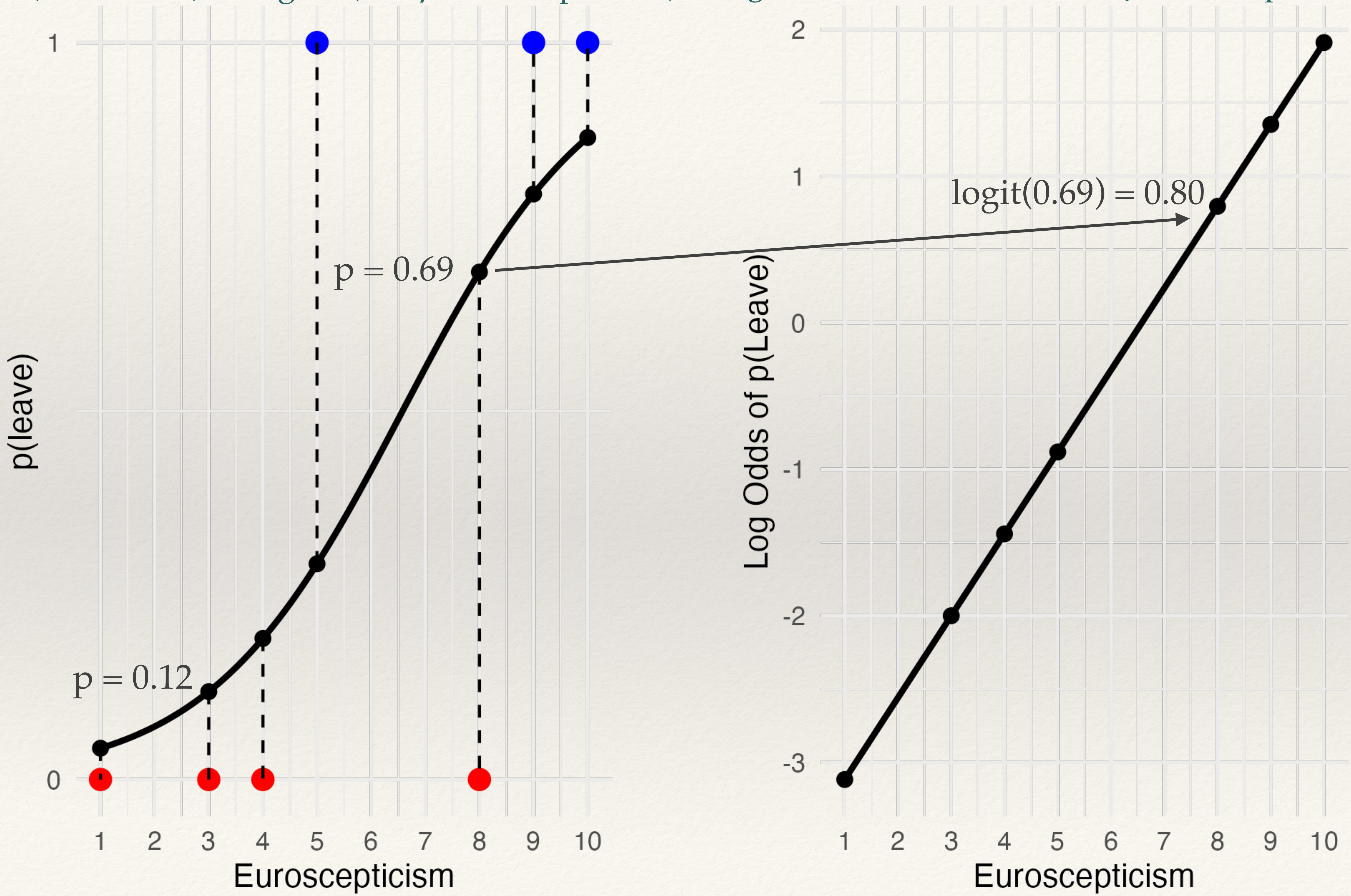


$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha + \beta \text{Euroscepticism})}}$$

$$\log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta \text{Euroscepticism}$$

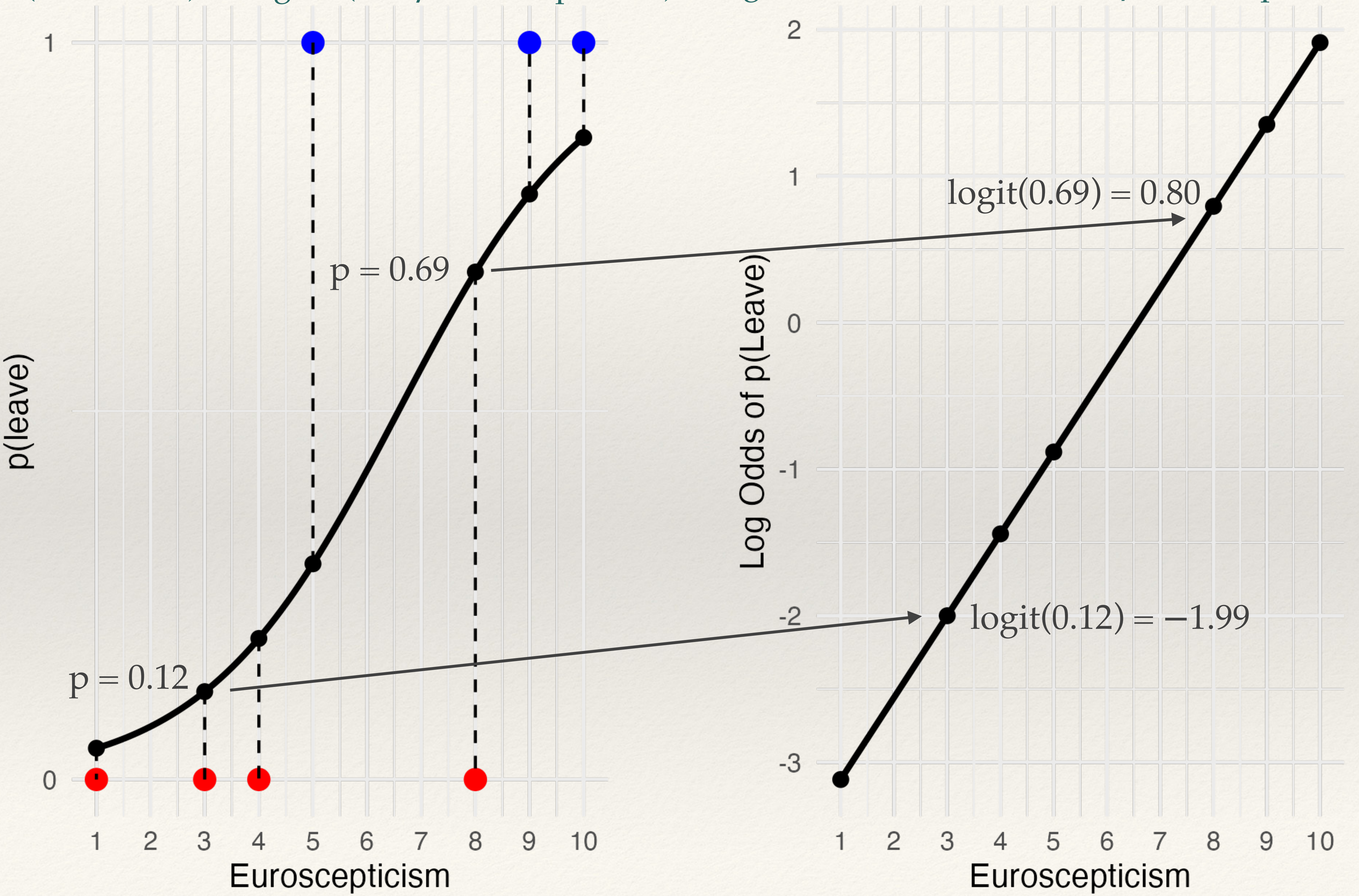
$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta \text{Euroscepticism})$$

$$\text{logit}[\Pr(\text{Leave} = 1)] = \alpha + \beta \text{Euroscepticism}$$



$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha + \beta \text{Euroscepticism})}} \qquad \log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta \text{Euroscepticism}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta \text{Euroscepticism}) \qquad \text{logit}[\Pr(\text{Leave} = 1)] = \alpha + \beta \text{Euroscepticism}$$

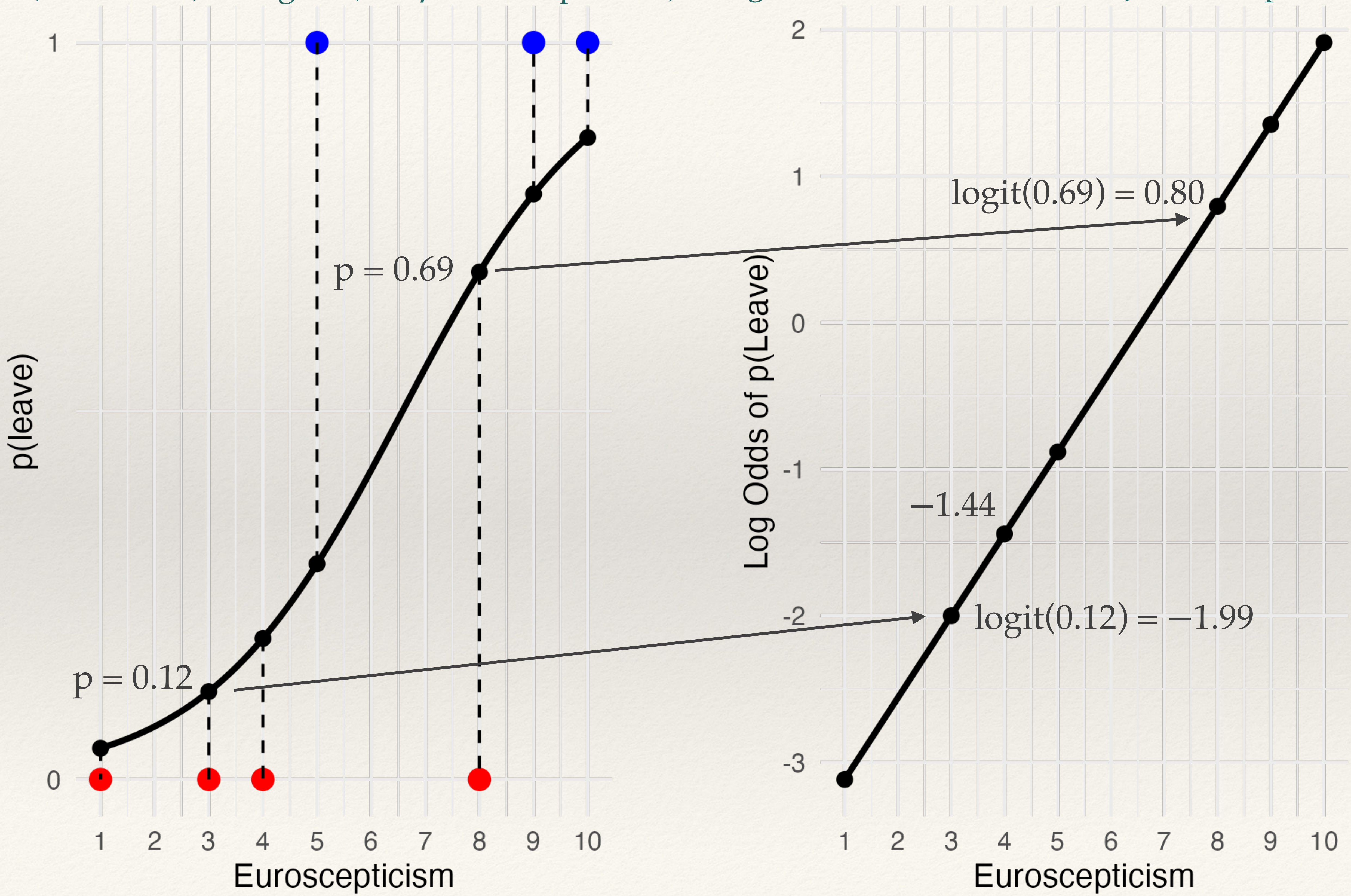


$$\Pr(\text{Leave} = 1) = \frac{1}{1 + e^{-(\alpha+\beta\text{Euroscepticism})}}$$

$$\log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta\text{Euroscepticism}$$

$$\Pr(\text{Leave} = 1) = \text{logit}^{-1}(\alpha + \beta\text{Euroscepticism})$$

$$\text{logit}[\Pr(\text{Leave} = 1)] = \alpha + \beta\text{Euroscepticism}$$

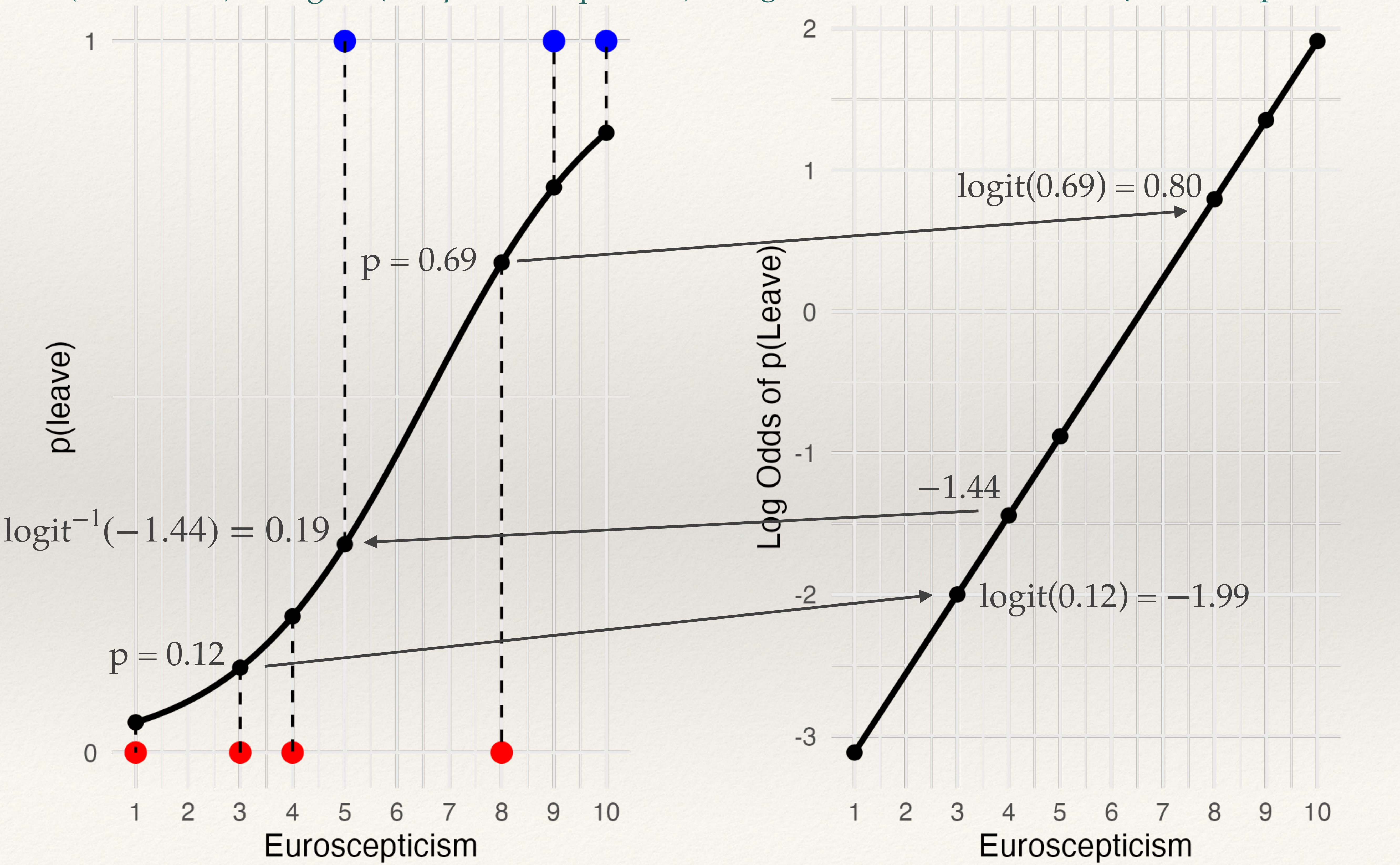


$$\Pr(\text{Leave}=1) = \frac{1}{1+e^{-(\alpha+\beta \text{Euroscepticism})}}$$

$$\log \frac{\Pr(\text{Leave}=1)}{\Pr(\text{Leave}=0)} = \alpha + \beta \text{Euroscepticism}$$

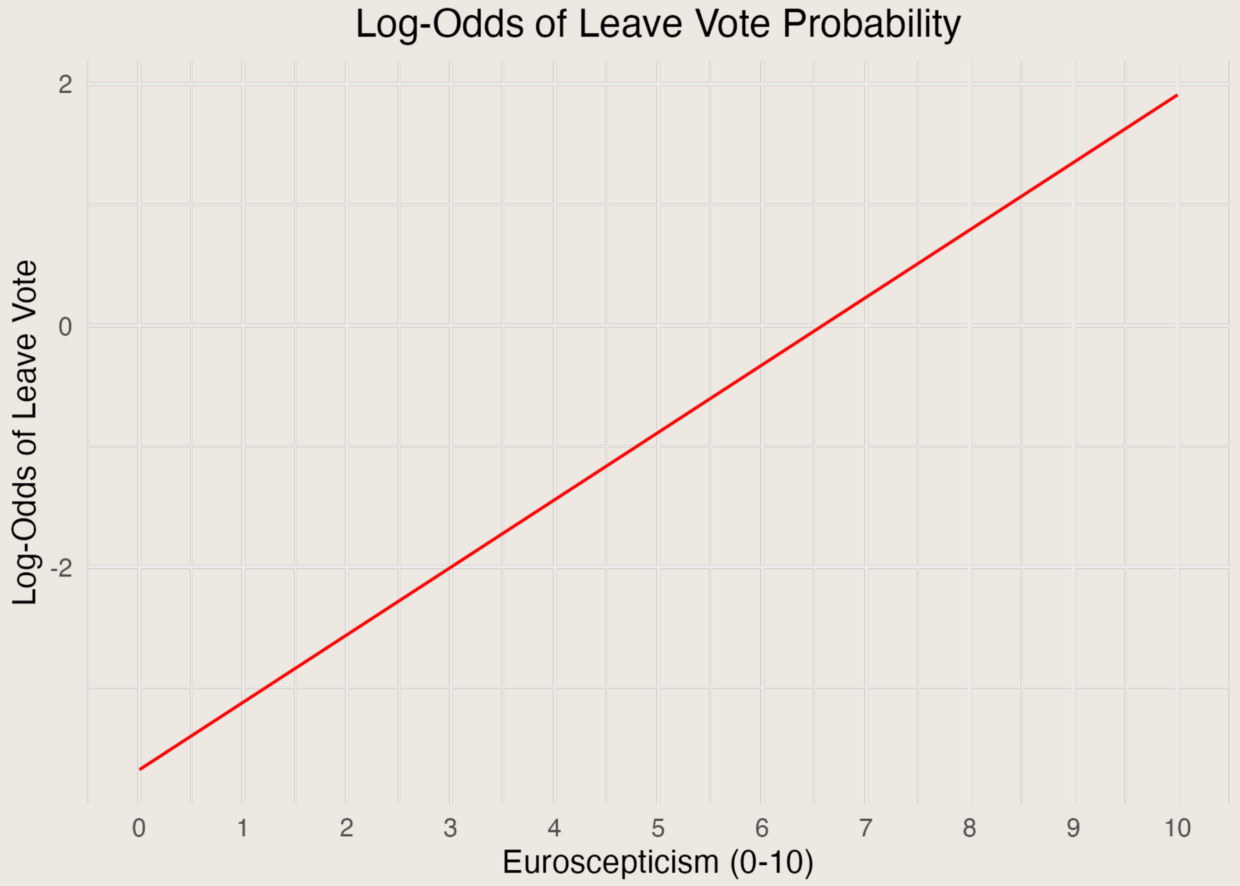
$$\Pr(\text{Leave}=1) = \text{logit}^{-1}(\alpha + \beta \text{Euroscepticism})$$

$$\text{logit}[\Pr(\text{Leave}=1)] = \alpha + \beta \text{Euroscepticism}$$



Logistic Regression Coefficients

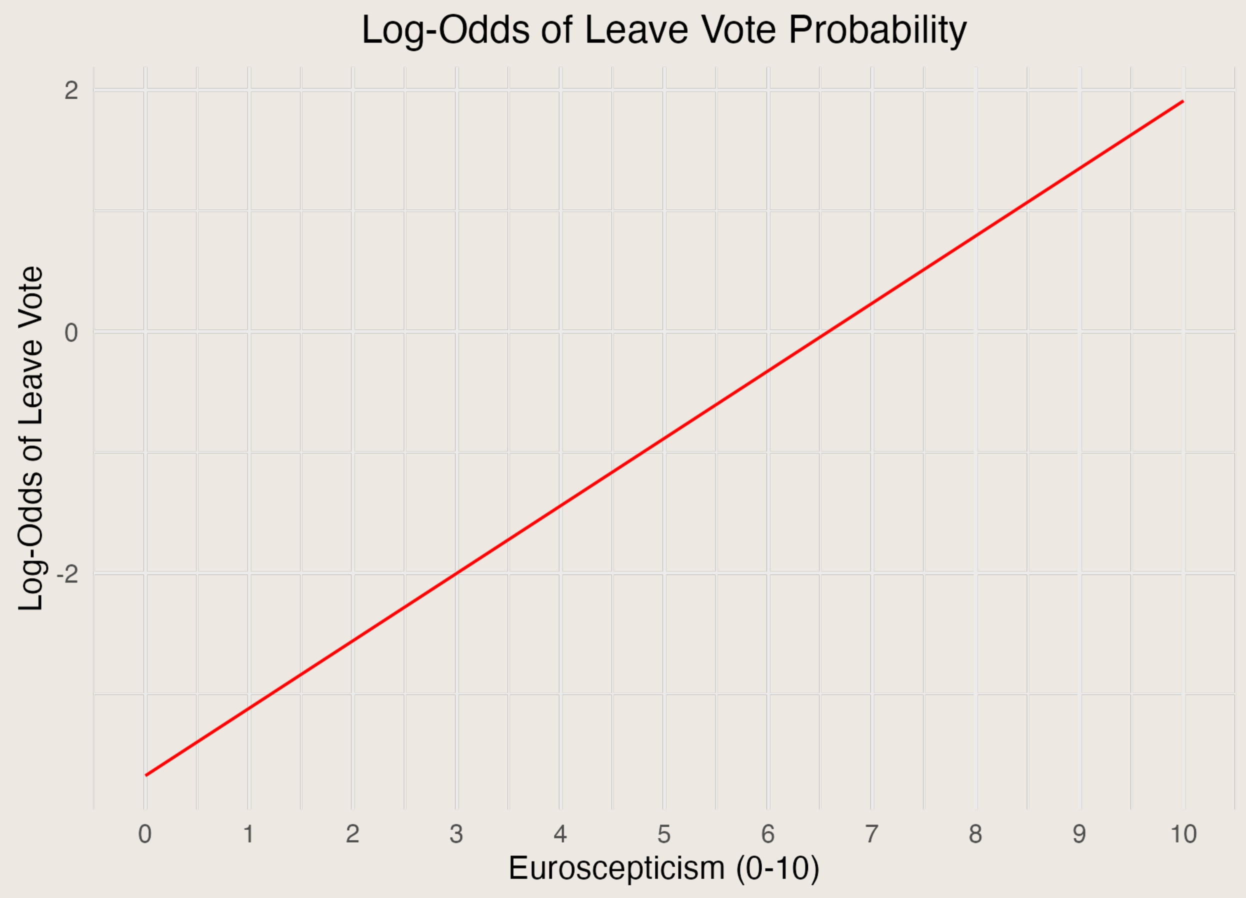
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

- Intercept: Log odds when X is zero: -3.68

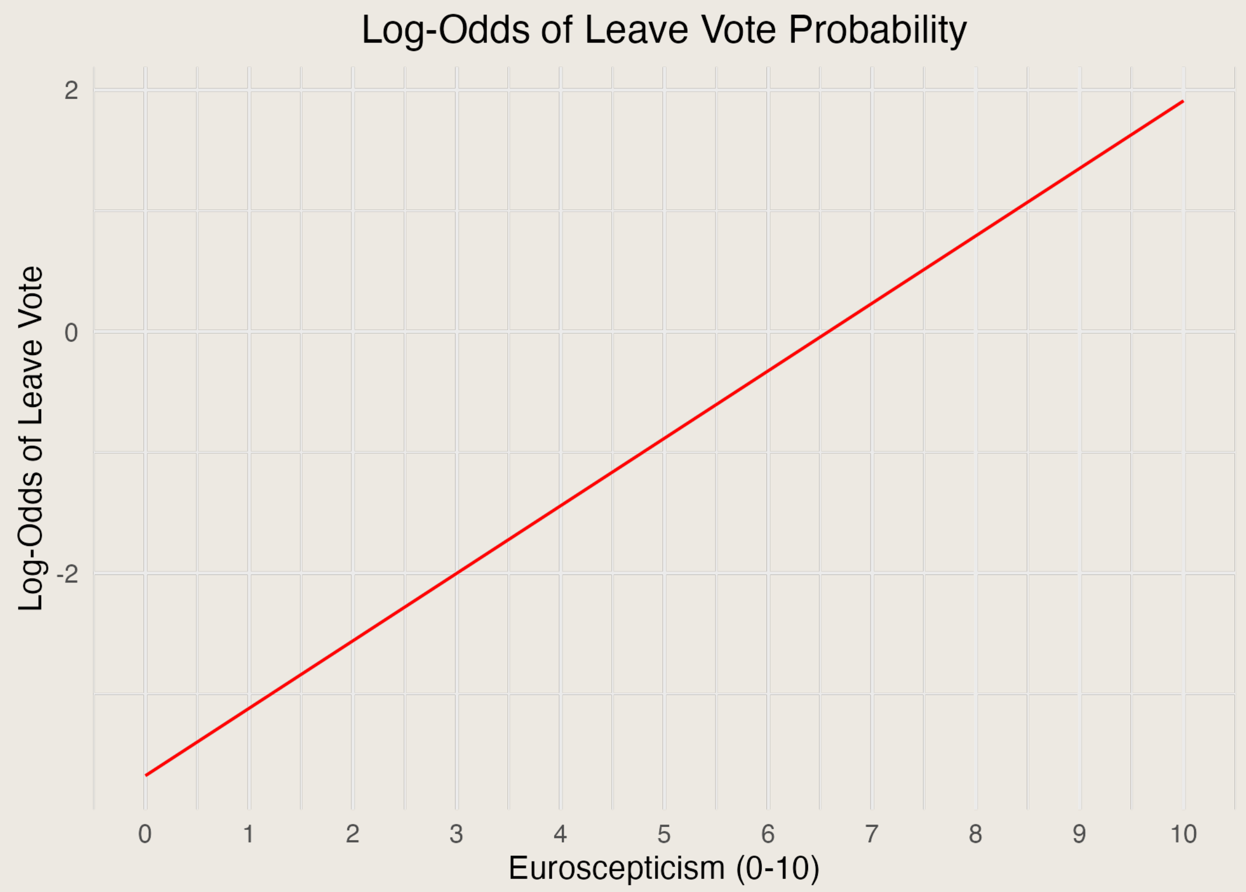
	Dependent variable:
	Leave Vote
Intercept	-3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

- Intercept: Log odds when X is zero: -3.68
- Slope: Predicted change in log-odds associated with a one-unit increase in X.

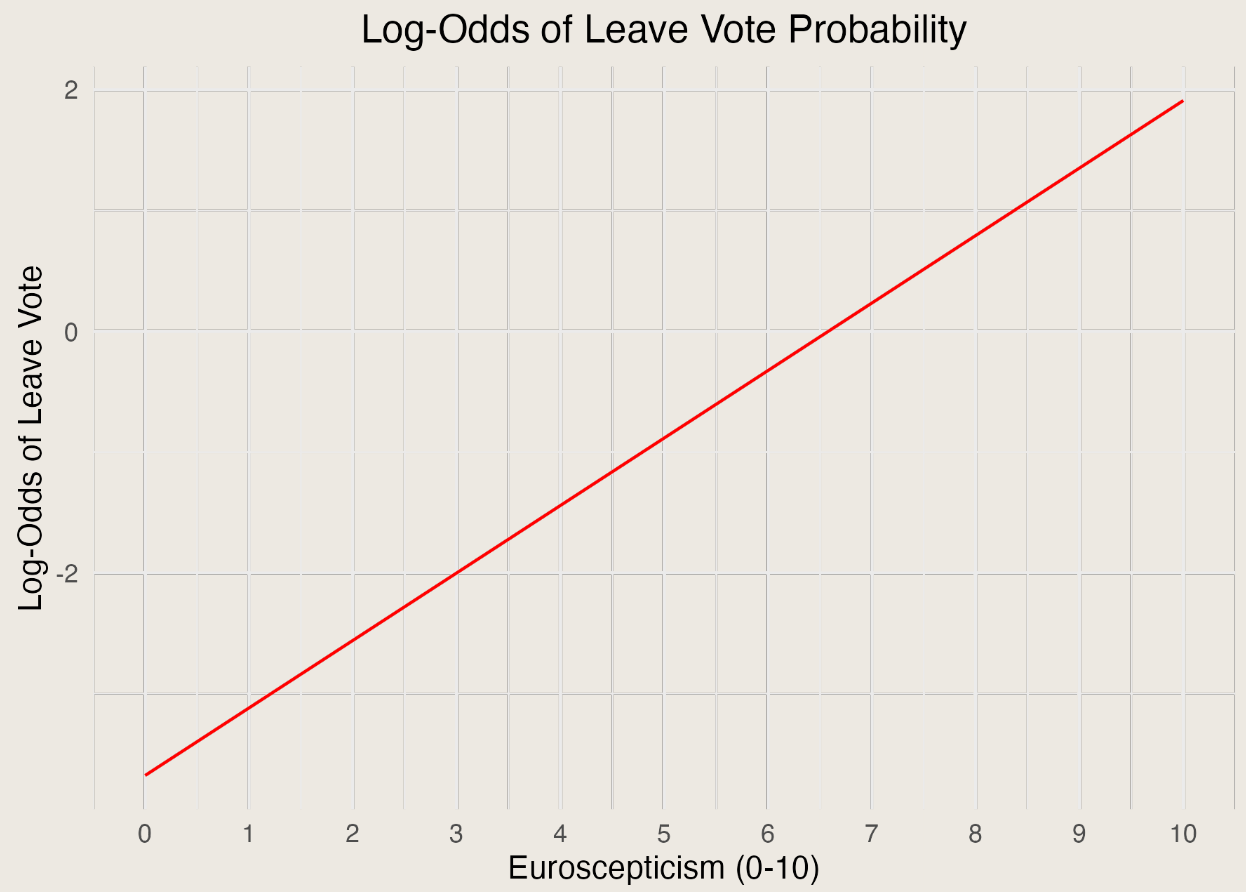
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

- Intercept: Log odds when X is zero: -3.68
- Slope: Predicted change in log-odds associated with a one-unit increase in X.
- **Log-odds of Leave vote when Euroscepticism = 0**: -3.68

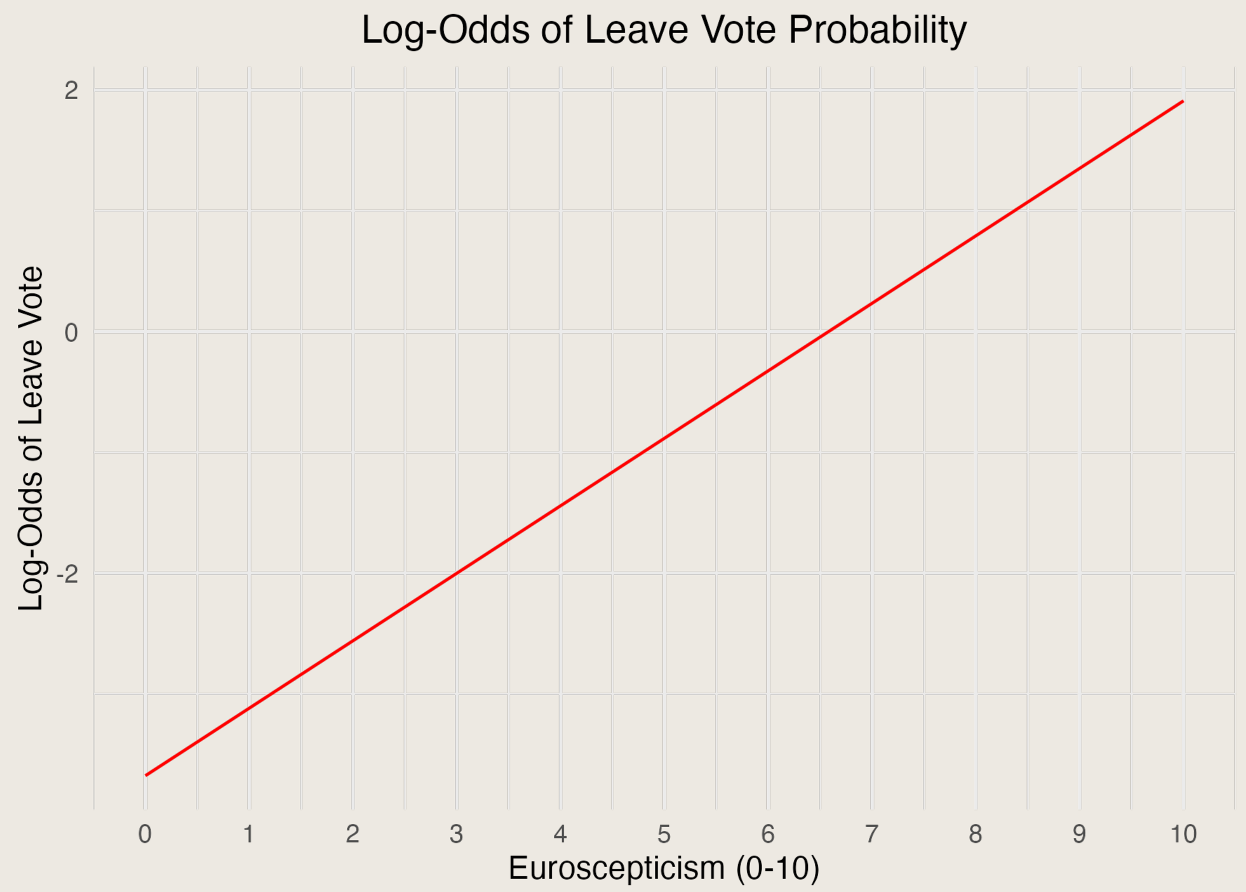
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

- Intercept: Log odds when X is zero: -3.68
- Slope: Predicted change in log-odds associated with a one-unit increase in X.
- **Log-odds of Leave vote when Euroscepticism = 0:** -3.68
- **Log-odds of Leave vote when Euroscepticism = 1:** $-3.68 + 0.56 = -3.12$

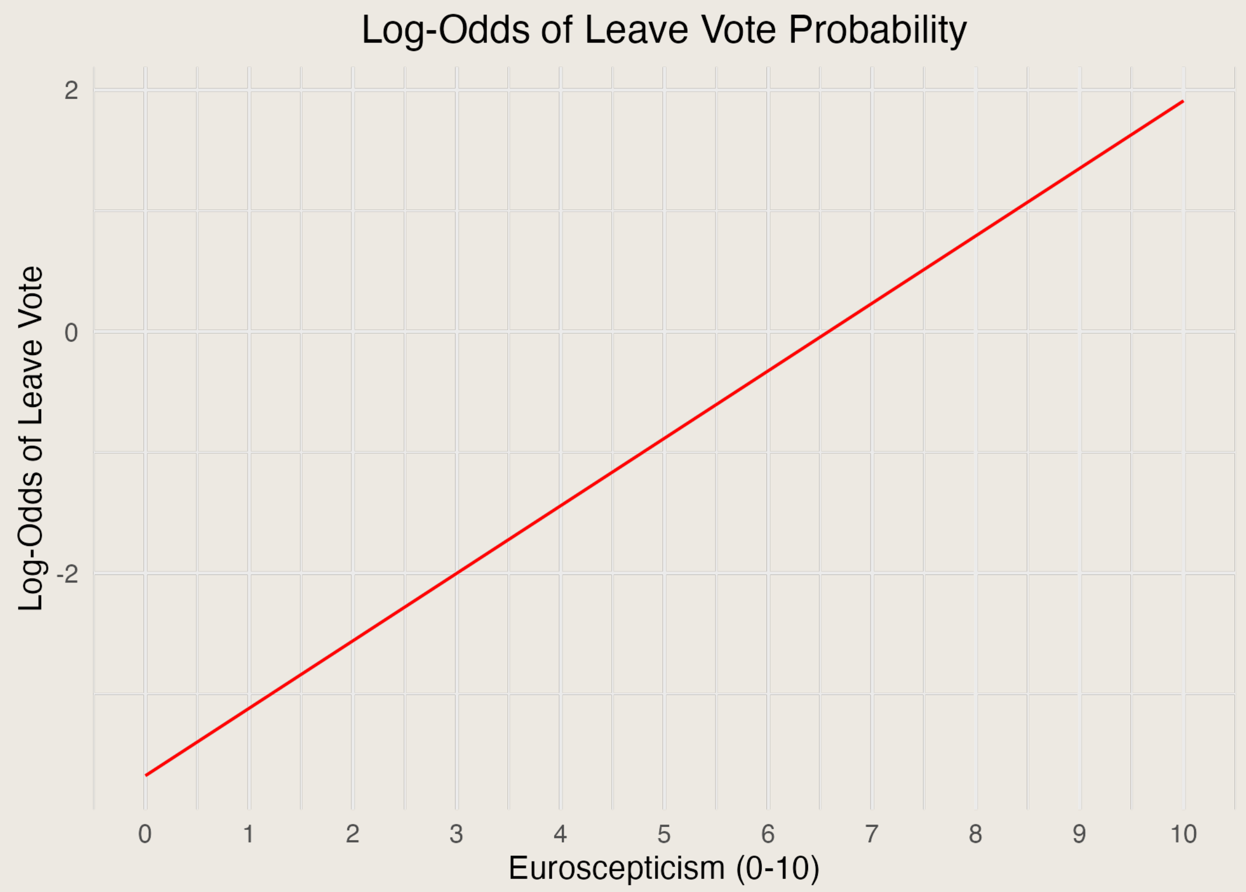
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

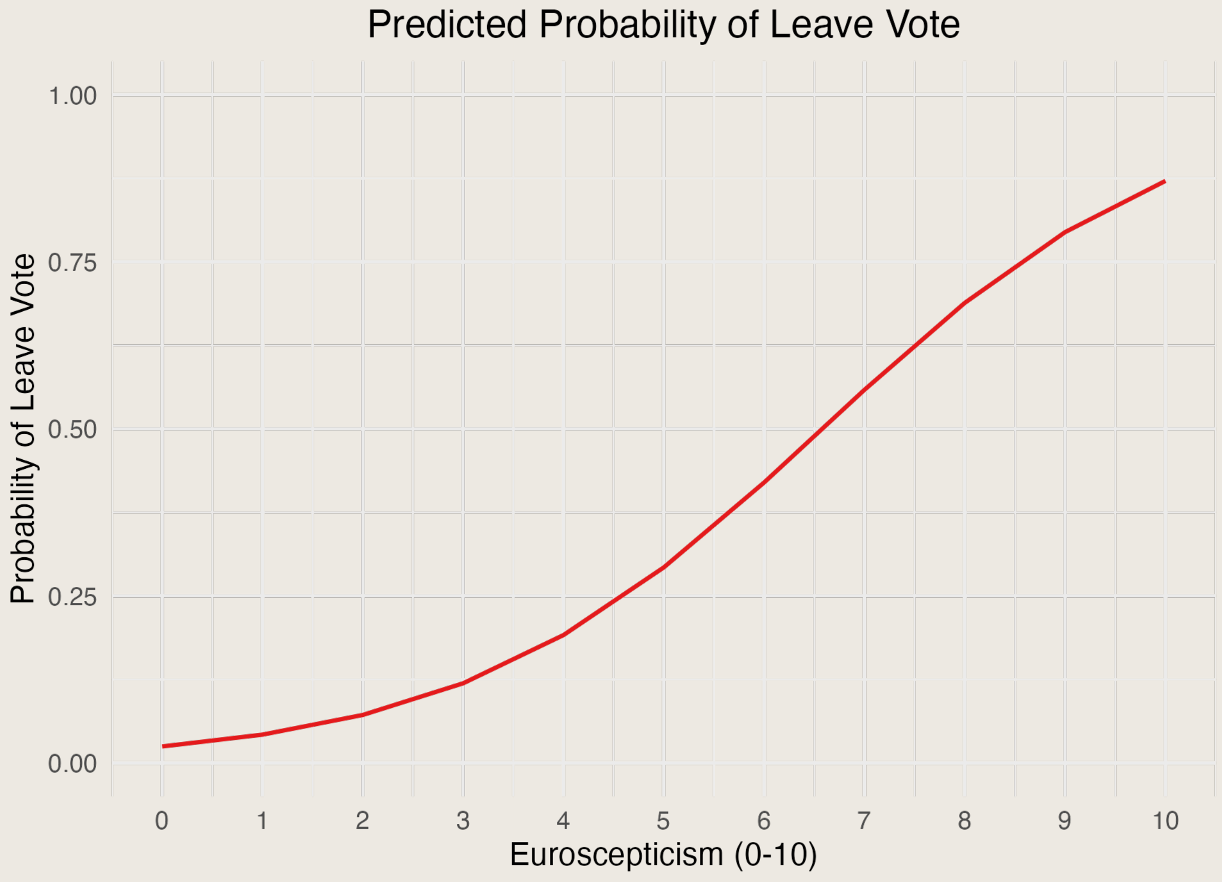
- Intercept: Log odds when X is zero: -3.68
- Slope: Predicted change in log-odds associated with a one-unit increase in X.
- **Log-odds of Leave vote when Euroscepticism = 0**: -3.68
- **Log-odds of Leave vote when Euroscepticism = 1**: $-3.68 + 0.56 = -3.12$
- **Log-odds of Leave vote when Euroscepticism = 2**: $-3.68 + 2 \times 0.56 = -2.56$

	Dependent variable:
	Leave Vote
Intercept	-3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

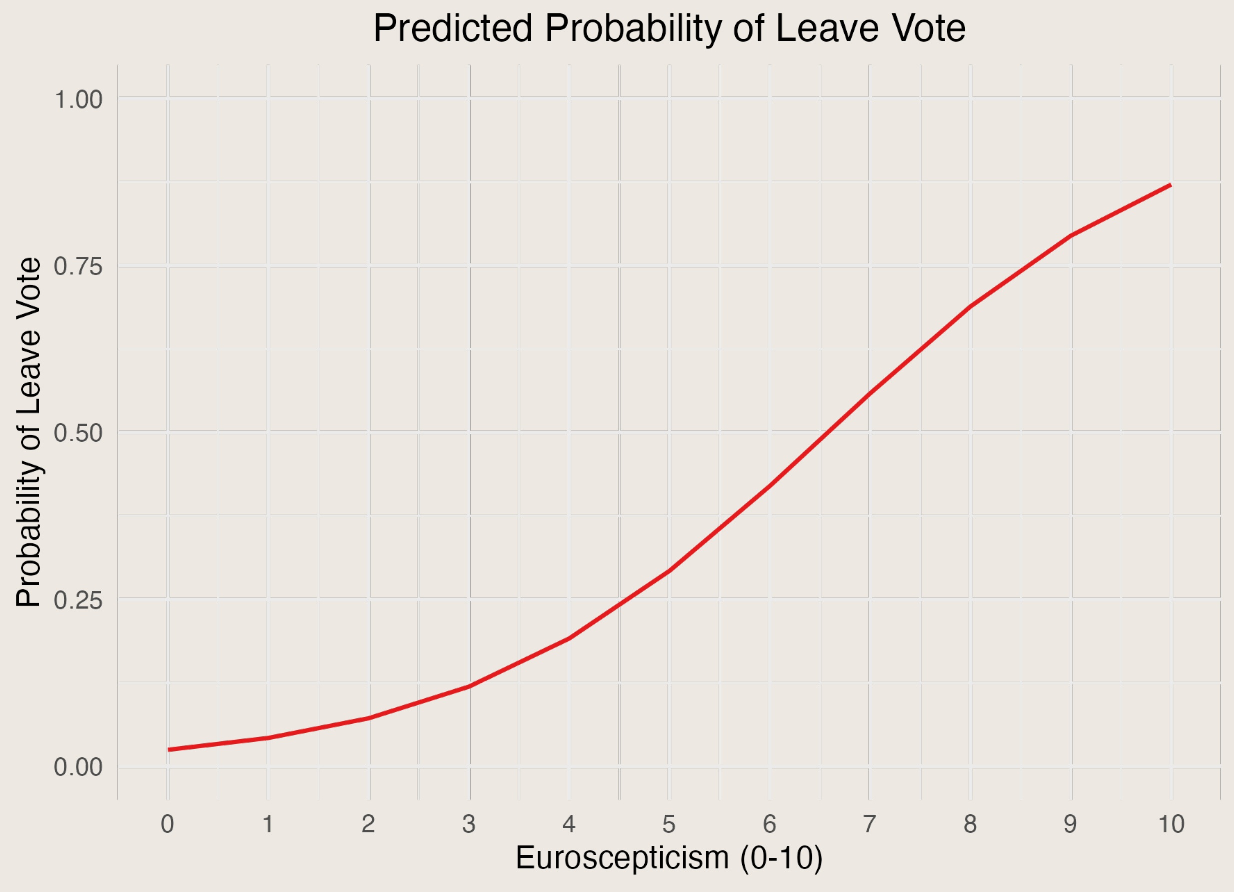
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

- Use **inverse-logit function** to get the predicted probability:

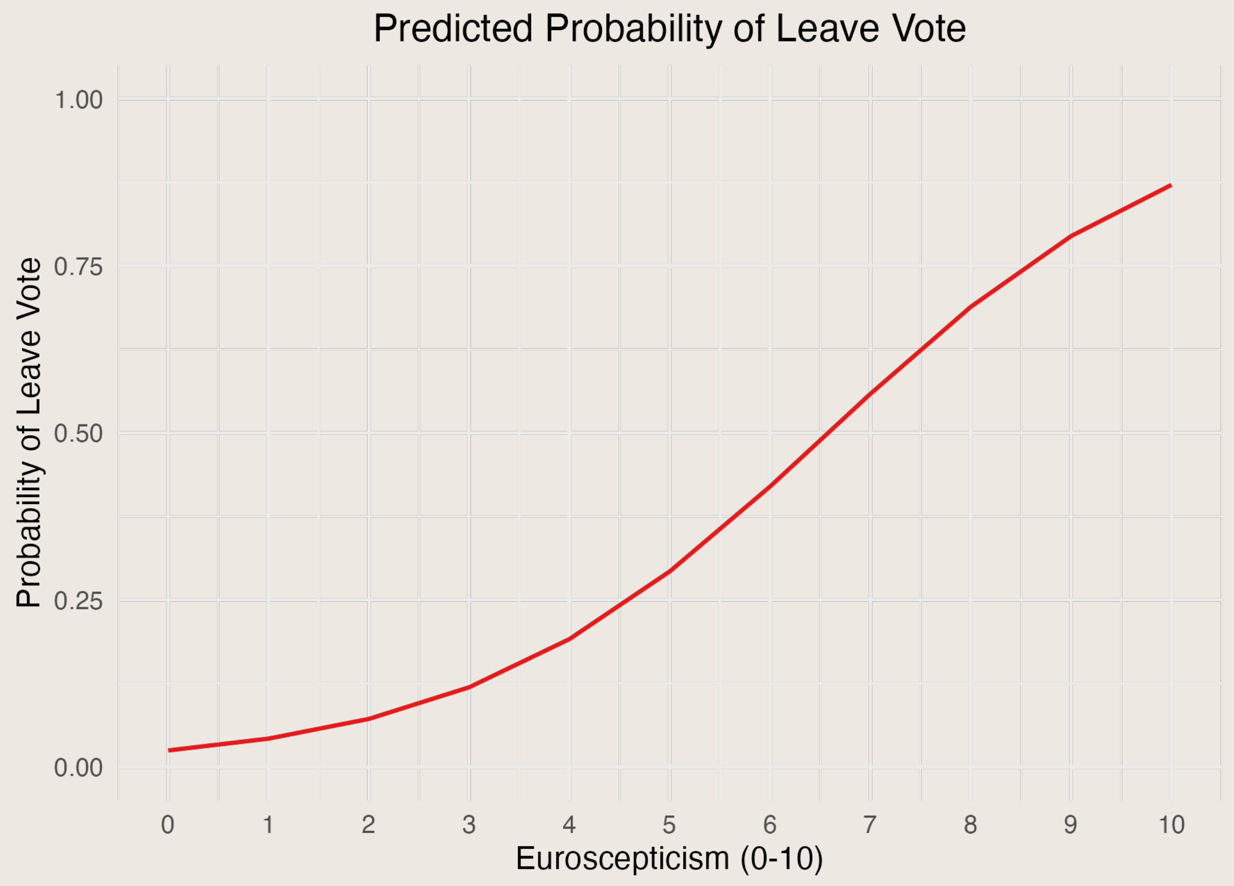
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

- Use **inverse-logit function** to get the predicted probability:
- **Probability of Leave vote for Euroscepticism = 0**
  $\text{logit}^{-1}(-3.68) = 0.024$

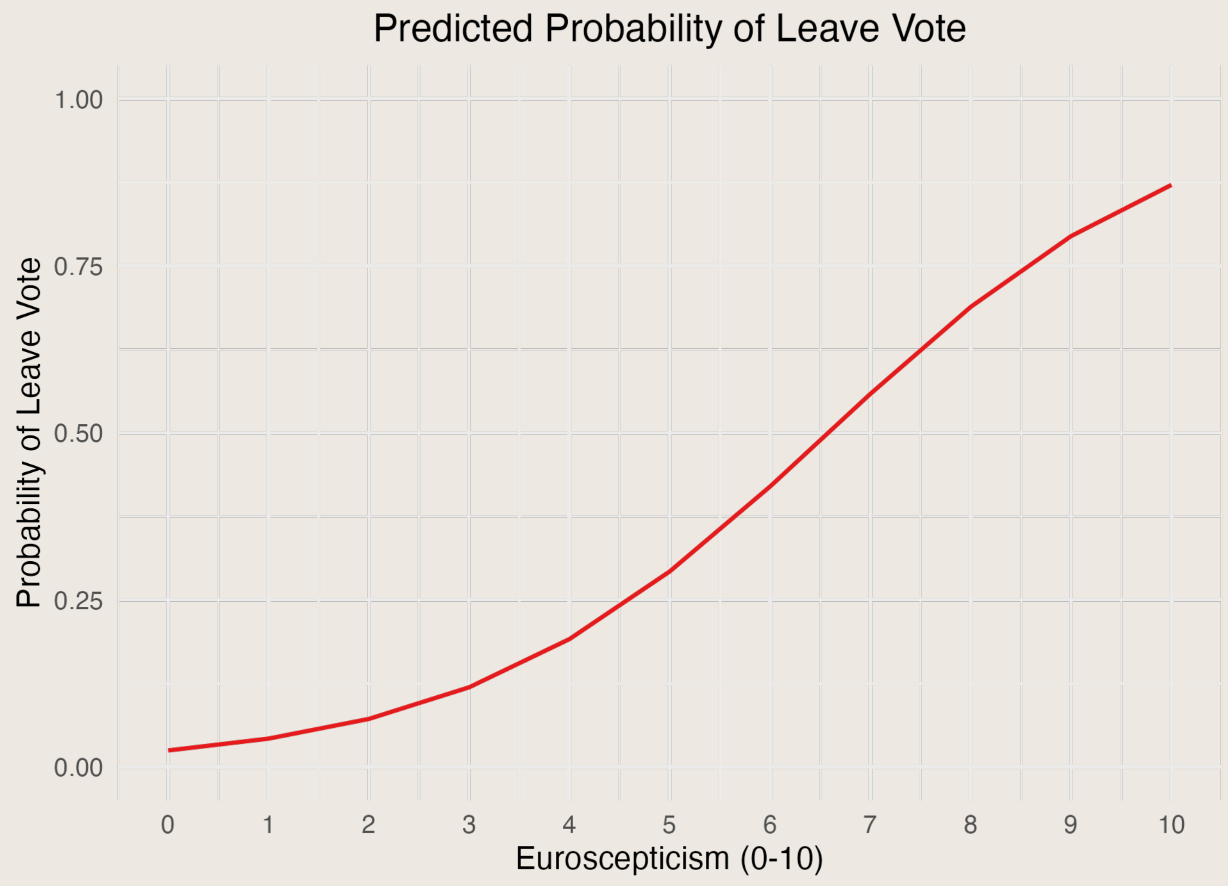
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

- Use **inverse-logit function** to get the predicted probability:
- **Probability of Leave vote for Euroscepticism = 0**
  $\text{logit}^{-1}(-3.68) = 0.024$
- **Probability of Leave vote for Euroscepticism = 1**
  $\text{logit}^{-1}(-3.68 + 0.56) = 0.042$

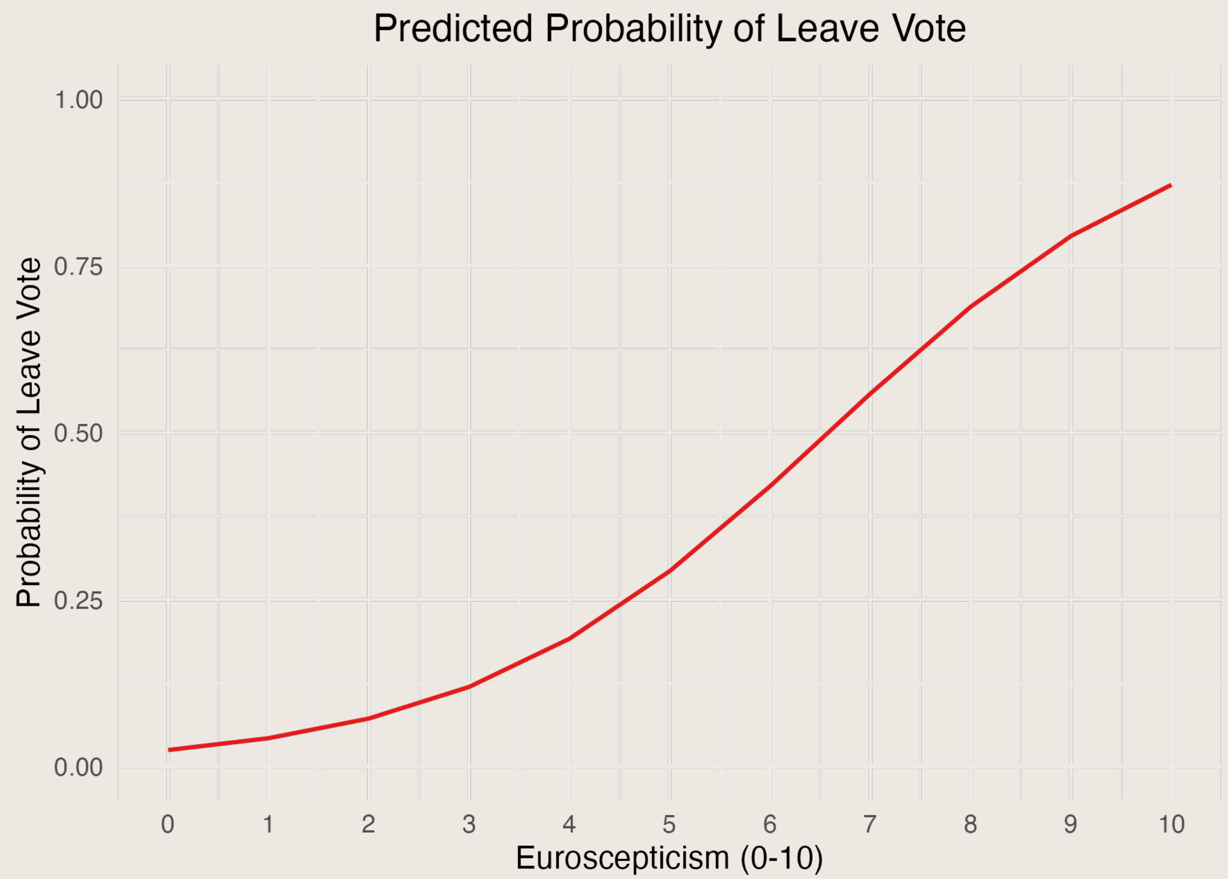
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

- Use **inverse-logit function** to get the predicted probability:
- **Probability of Leave vote for Euroscepticism = 0**
  $\text{logit}^{-1}(-3.68) = 0.024$
- **Probability of Leave vote for Euroscepticism = 1**
  $\text{logit}^{-1}(-3.68 + 0.56) = 0.042$
- **Probability of Leave vote for Euroscepticism = 2**
  $\text{logit}^{-1}(-3.68 + 2 \times 0.56) = 0.072$

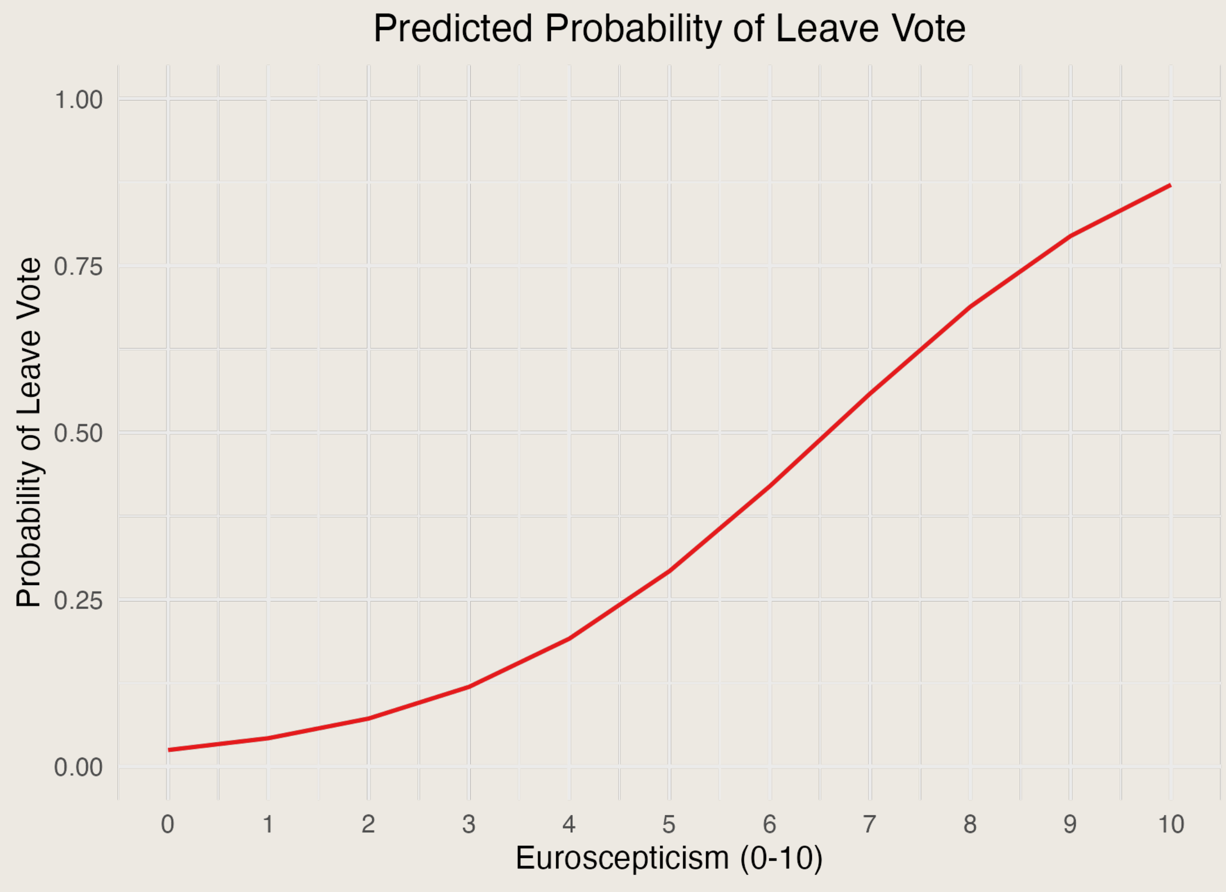
	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression Coefficients

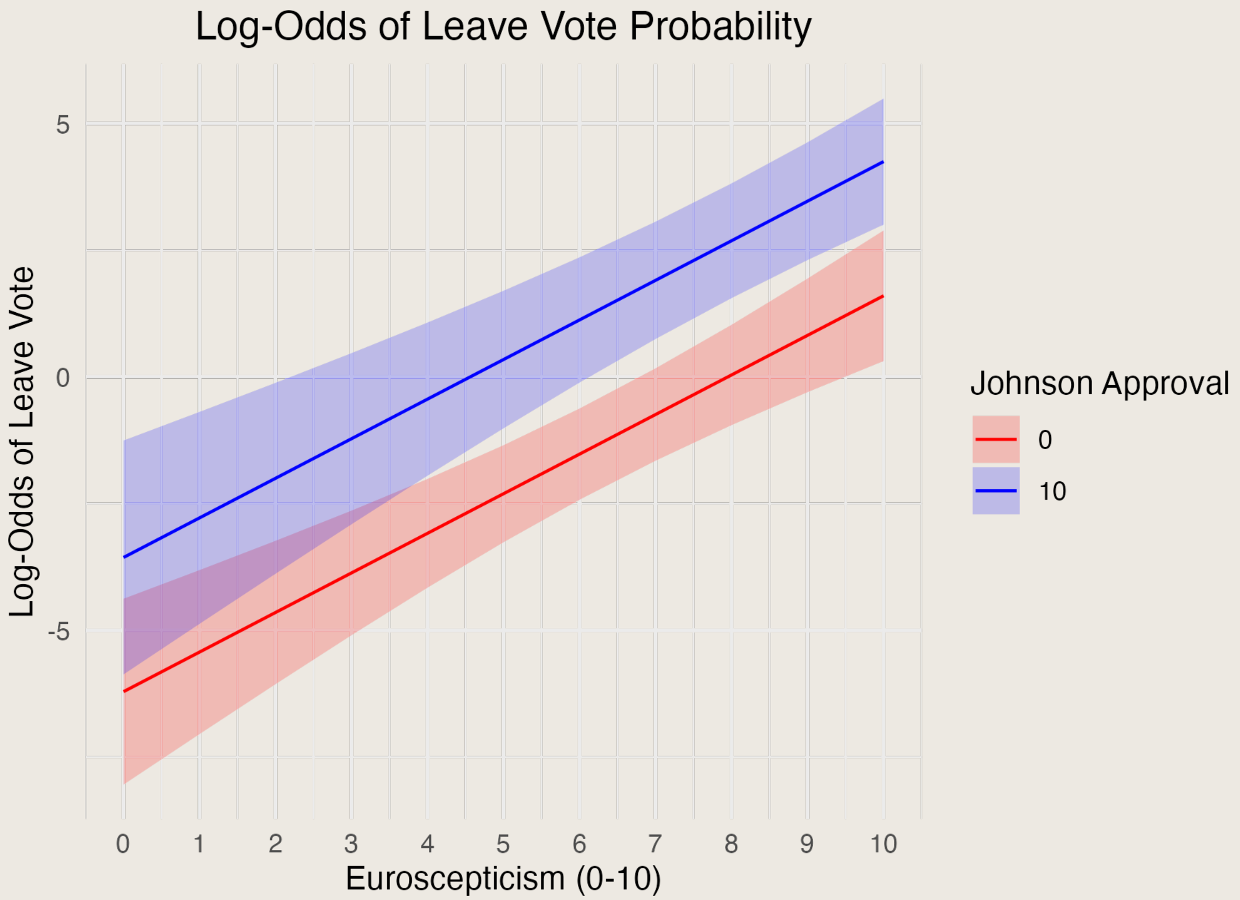
- Use **inverse-logit function** to get the predicted probability:
- **Probability of Leave vote for Euroscepticism = 0**
  $\text{logit}^{-1}(-3.68) = 0.024$
- **Probability of Leave vote for Euroscepticism = 1**
  $\text{logit}^{-1}(-3.68 + 0.56) = 0.042$
- **Probability of Leave vote for Euroscepticism = 2**
  $\text{logit}^{-1}(-3.68 + 2 \times 0.56) = 0.072$
- **Probability of Leave vote for Euroscepticism = 3**
  $\text{logit}^{-1}(-3.68 + 3 \times 0.56) = 0.12$

	Dependent variable:
	Leave Vote
Intercept	−3.68 (2.63)
Euroscepticism	0.56 (0.38)
Observations	7



Logistic Regression: Multiple Predictors

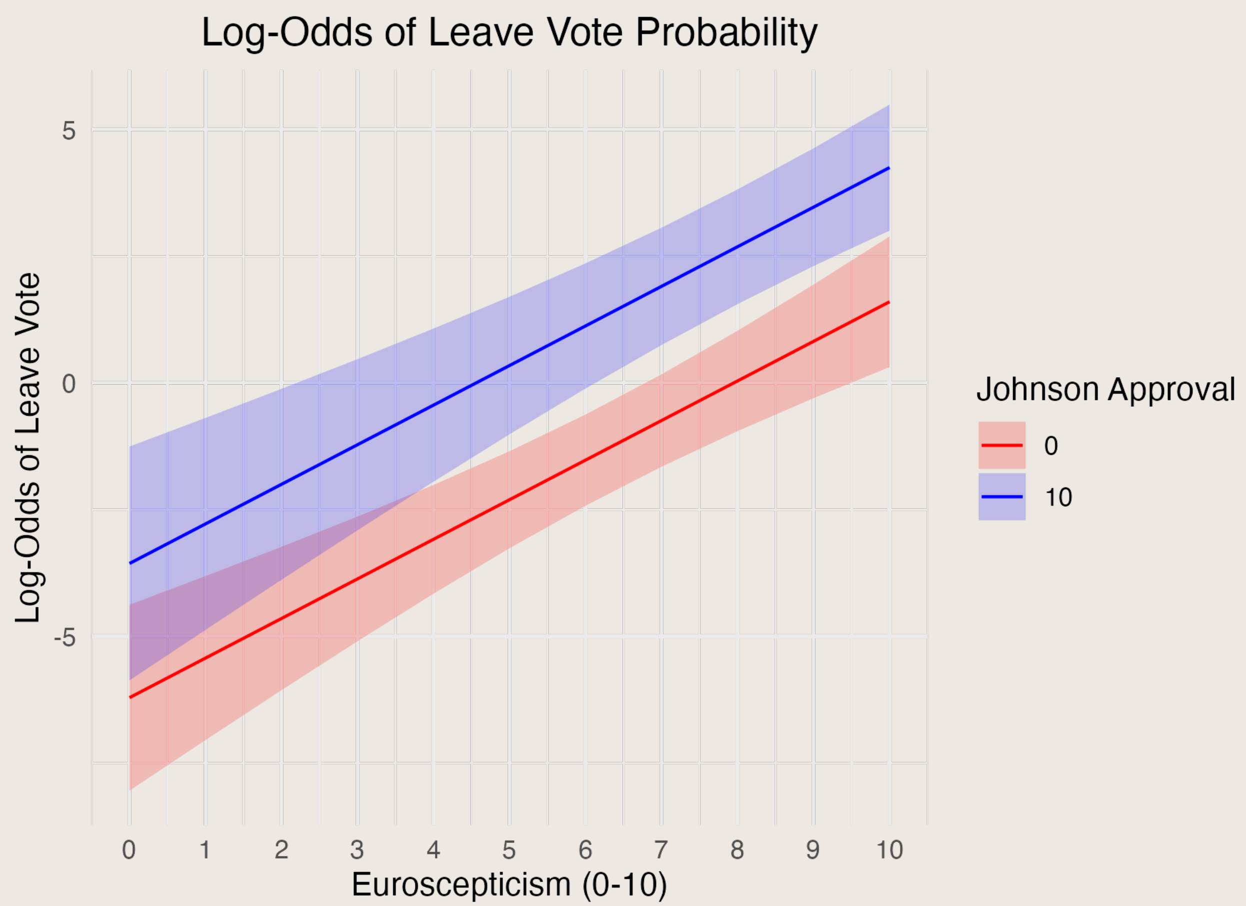
	Dependent
	Leave Vote
Intercept	−6.21 (0.93)
Euroscepticism	0.78 (0.13)
Johnson Approval	0.26 (0.09)
Observations	200



Logistic Regression: Multiple Predictors

- With multiple predictors, the **change in log-odds** associated with each predictor is still **linear**.

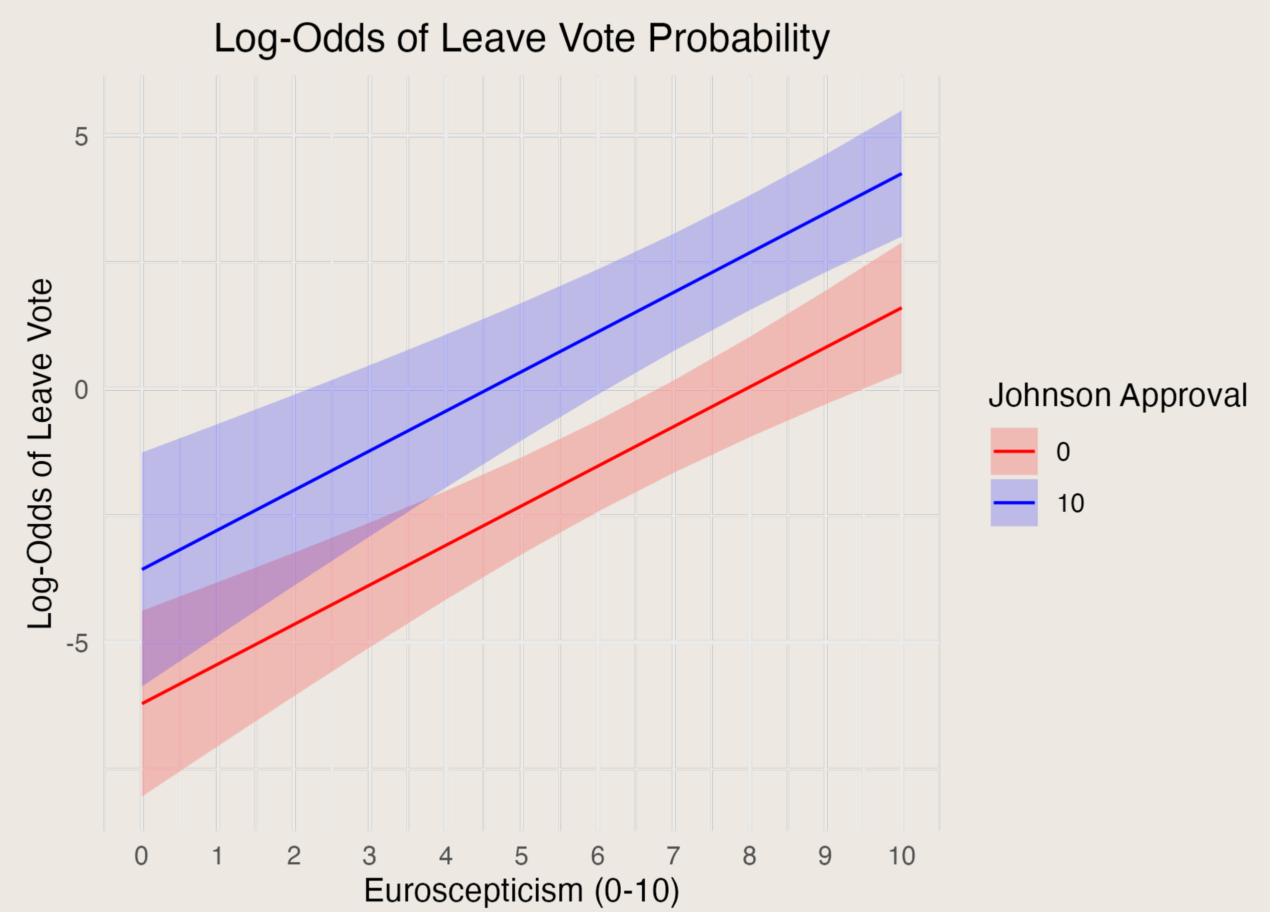
	Dependent
	Leave Vote
Intercept	−6.21 (0.93)
Euroscepticism	0.78 (0.13)
Johnson Approval	0.26 (0.09)
Observations	200



Logistic Regression: Multiple Predictors

- With multiple predictors, the **change in log-odds** associated with each predictor is still **linear**.
- The log-odds of Leave vote probability for someone who scores '0' on Euroscepticism and '0' on Johnson Approval is −6.21.

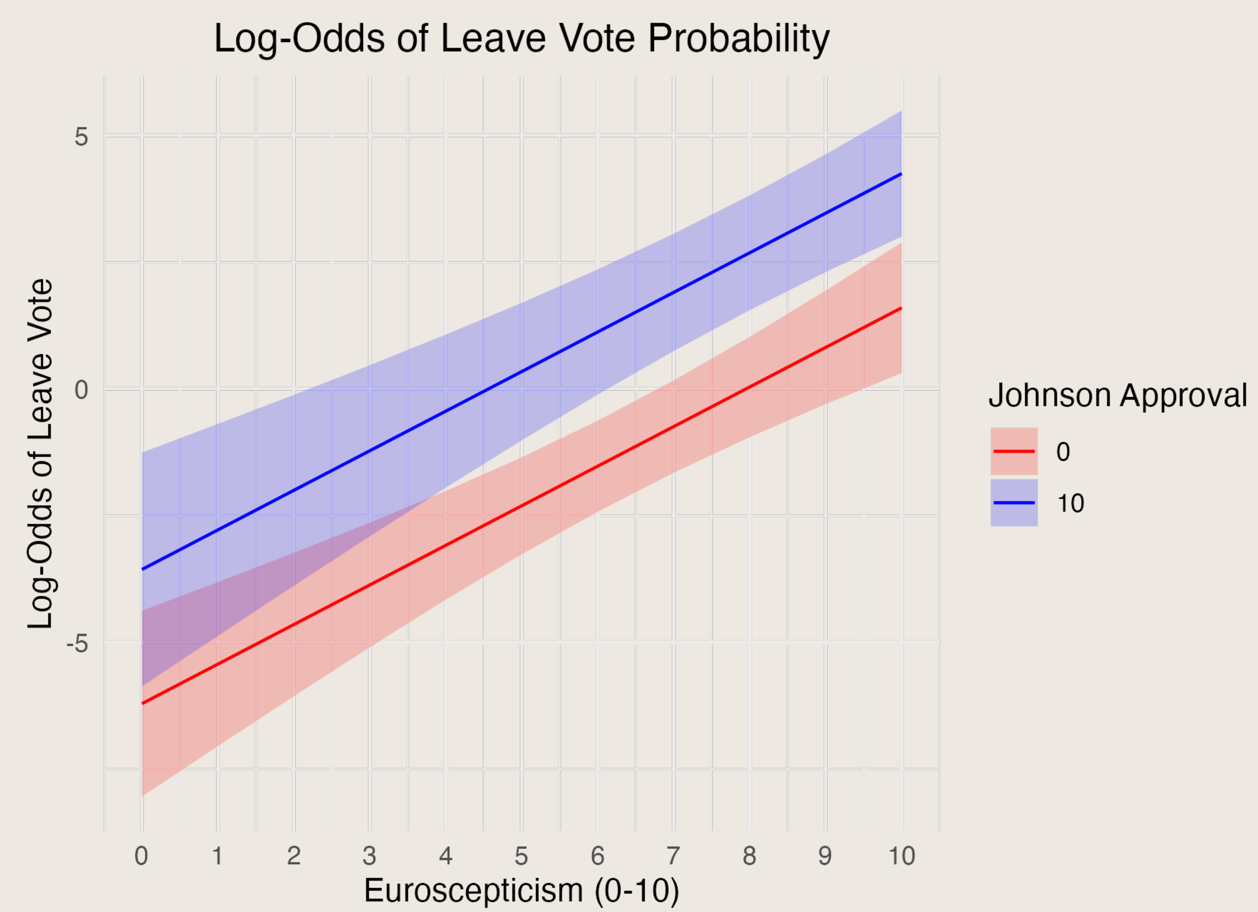
	Dependent
	Leave Vote
Intercept	−6.21 (0.93)
Euroscepticism	0.78 (0.13)
Johnson Approval	0.26 (0.09)
Observations	200



Logistic Regression: Multiple Predictors

- With multiple predictors, the **change in log-odds** associated with each predictor is still **linear**.
- The log-odds of Leave vote probability for someone who scores '0' on Euroscepticism and '0' on Johnson Approval is −6.21.
- For each one-point increase in Euroscepticism, the predicted log-odds increase by 0.78.

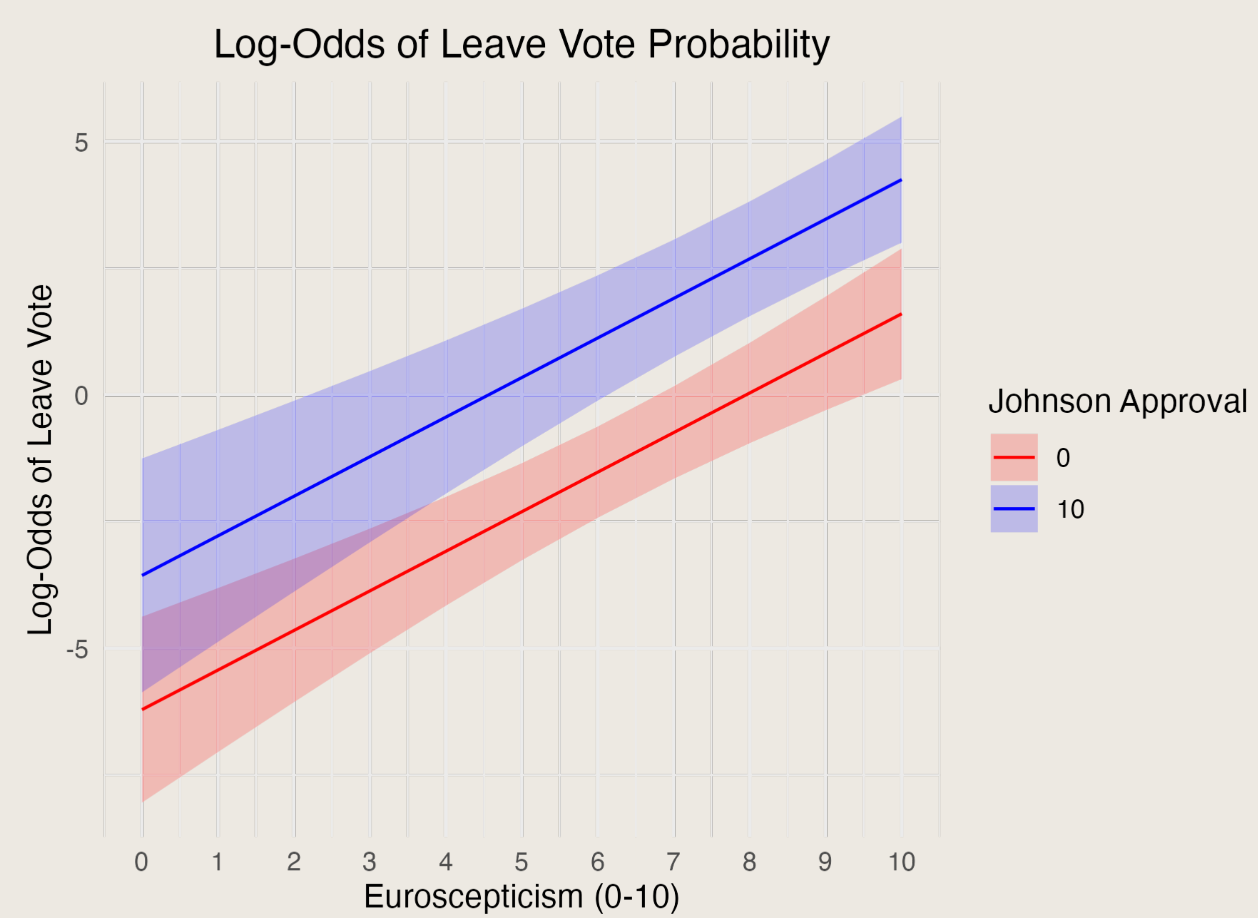
	Dependent
	Leave Vote
Intercept	−6.21 (0.93)
Euroscepticism	0.78 (0.13)
Johnson Approval	0.26 (0.09)
Observations	200



Logistic Regression: Multiple Predictors

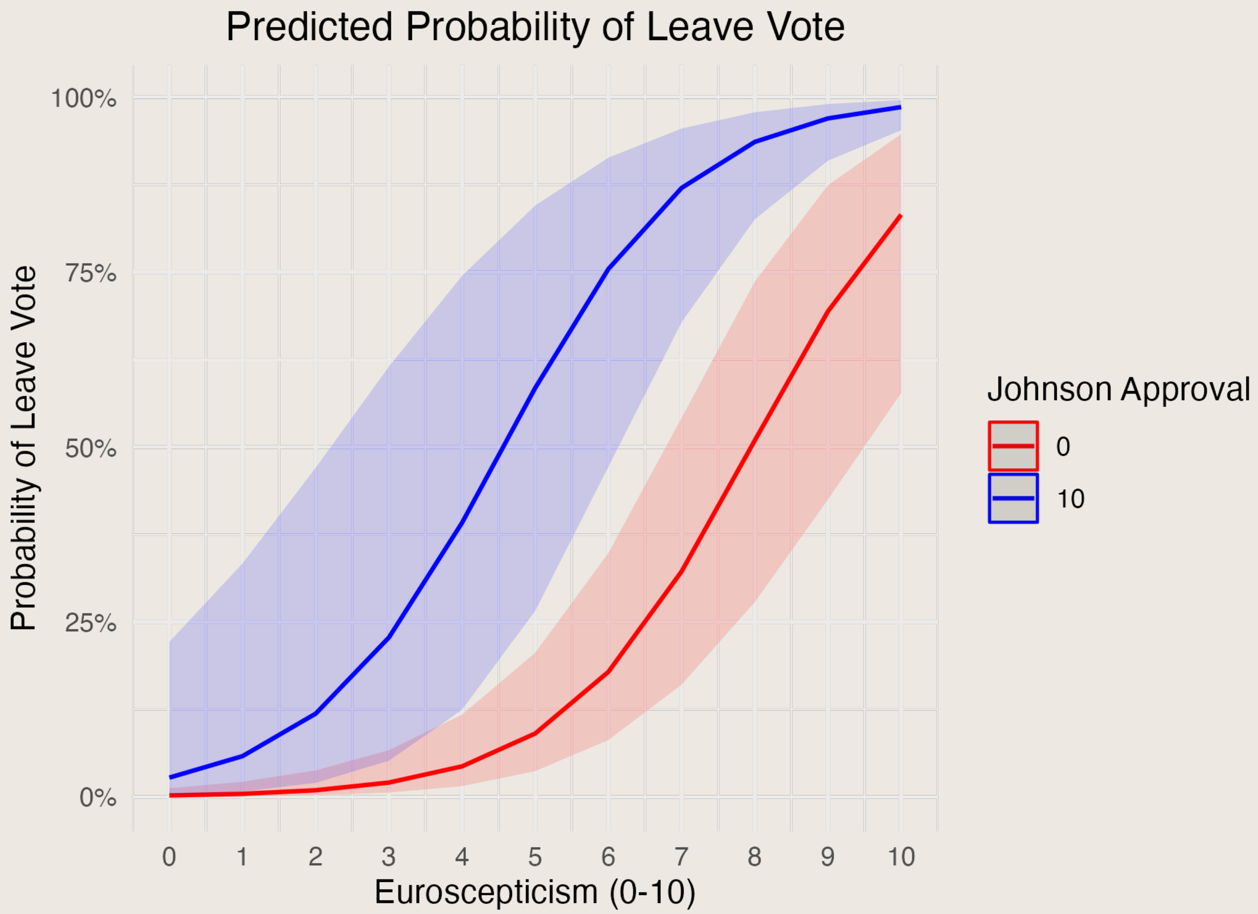
- With multiple predictors, the **change in log-odds** associated with each predictor is still **linear**.
- The log-odds of Leave vote probability for someone who scores '0' on Euroscepticism and '0' on Johnson Approval is −6.21.
- For each one-point increase in Euroscepticism, the predicted log-odds increase by 0.78.
- For each one-point increase in Johnson Approval, the predicted log-odds increase by 0.26.

	Dependent
	Leave Vote
Intercept	−6.21 (0.93)
Euroscepticism	0.78 (0.13)
Johnson Approval	0.26 (0.09)
Observations	200



Logistic Regression: Multiple Predictors

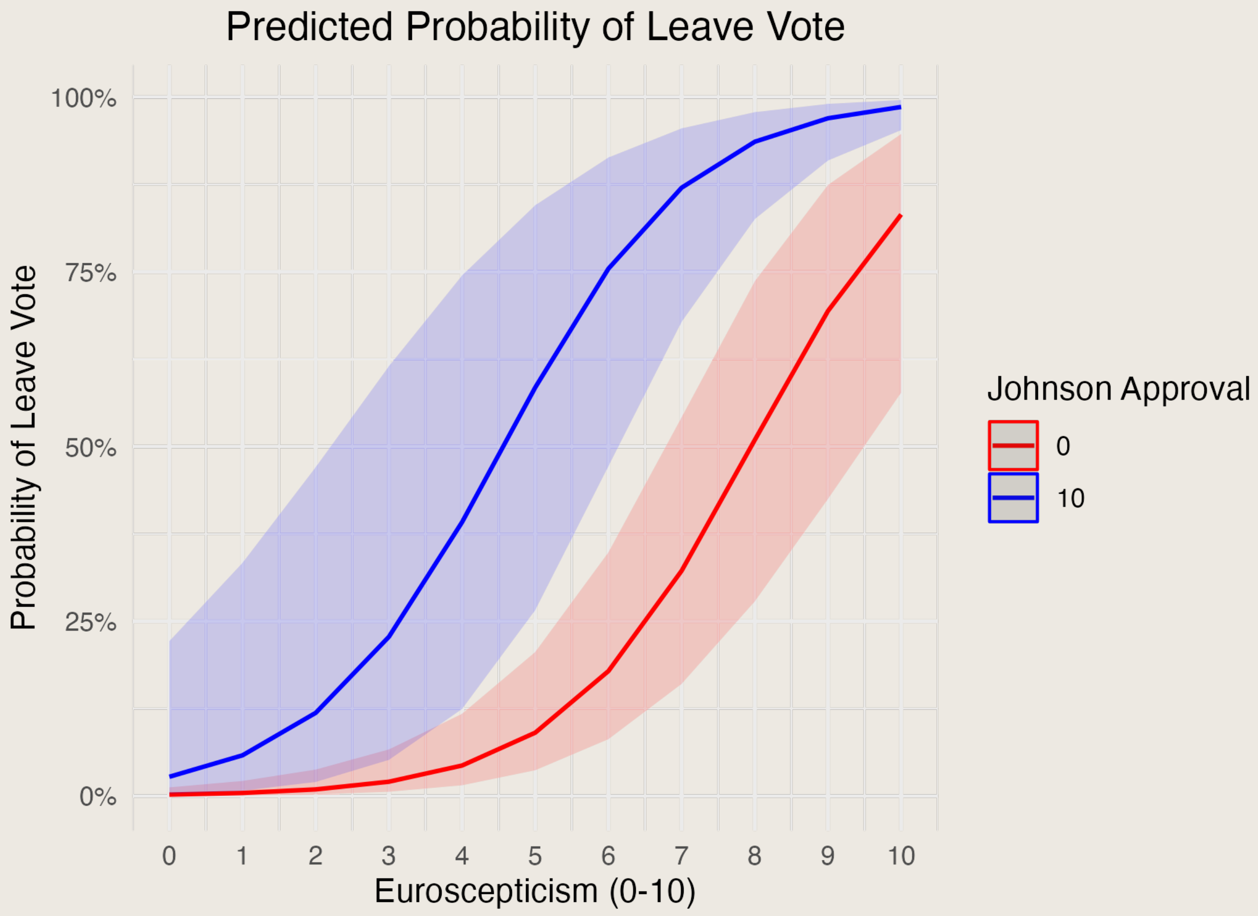
	Dependent
	Leave Vote
Intercept	-6.21^{***} (0.93)
Euroscepticism	0.78^{***} (0.13)
Johnson Approval	0.26^{***} (0.09)
Observations	200



Logistic Regression: Multiple Predictors

* Translating this into predicted probabilities is **trickier.**

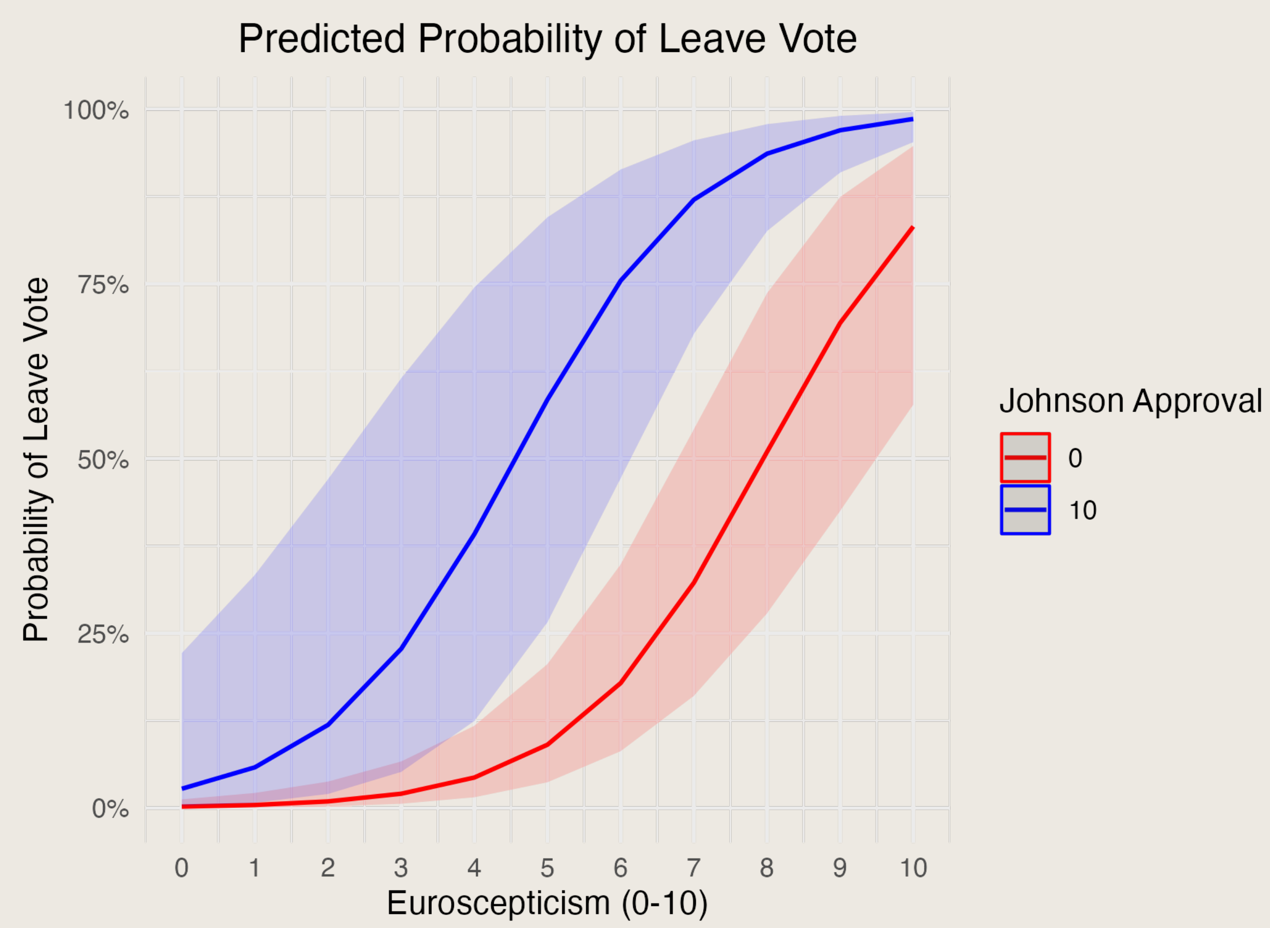
	Dependent
	Leave Vote
Intercept	-6.21^{***} (0.93)
Euroscepticism	0.78^{***} (0.13)
Johnson Approval	0.26^{***} (0.09)
Observations	200



Logistic Regression: Multiple Predictors

- Translating this into predicted probabilities is **trickier.**
- The predicted change in probability associated with a one-unit increase in Euroscepticism depends **both** on the level of Euroscepticism **and** on the level of Johnson Approval…

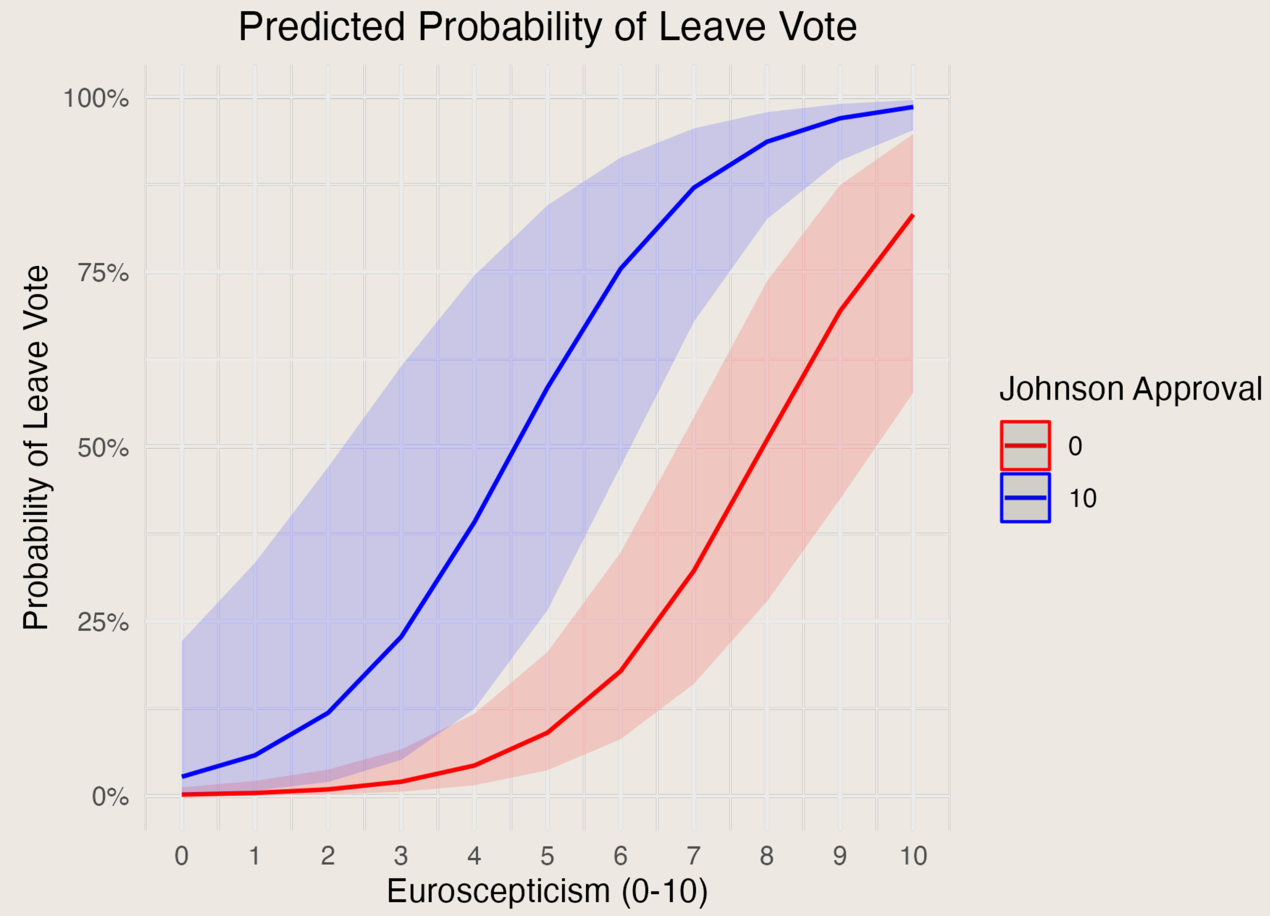
	Dependent
	Leave Vote
Intercept	-6.21^{***} (0.93)
Euroscepticism	0.78^{***} (0.13)
Johnson Approval	0.26^{***} (0.09)
Observations	200



Logistic Regression: Multiple Predictors

- Translating this into predicted probabilities is **trickier.**
- The predicted change in probability associated with a one-unit increase in Euroscepticism depends **both** on the level of Euroscepticism **and** on the level of Johnson Approval…
- In complex models, interpret sign and significance of coefficients, do not interpret their value.

	Dependent
	Leave Vote
Intercept	-6.21^{***} (0.93)
Euroscepticism	0.78^{***} (0.13)
Johnson Approval	0.26^{***} (0.09)
Observations	200



Visualising Predicted Probabilities

Visualising Predicted Probabilities

- Better (**but not best**) approach: predicted probabilities.

Visualising Predicted Probabilities

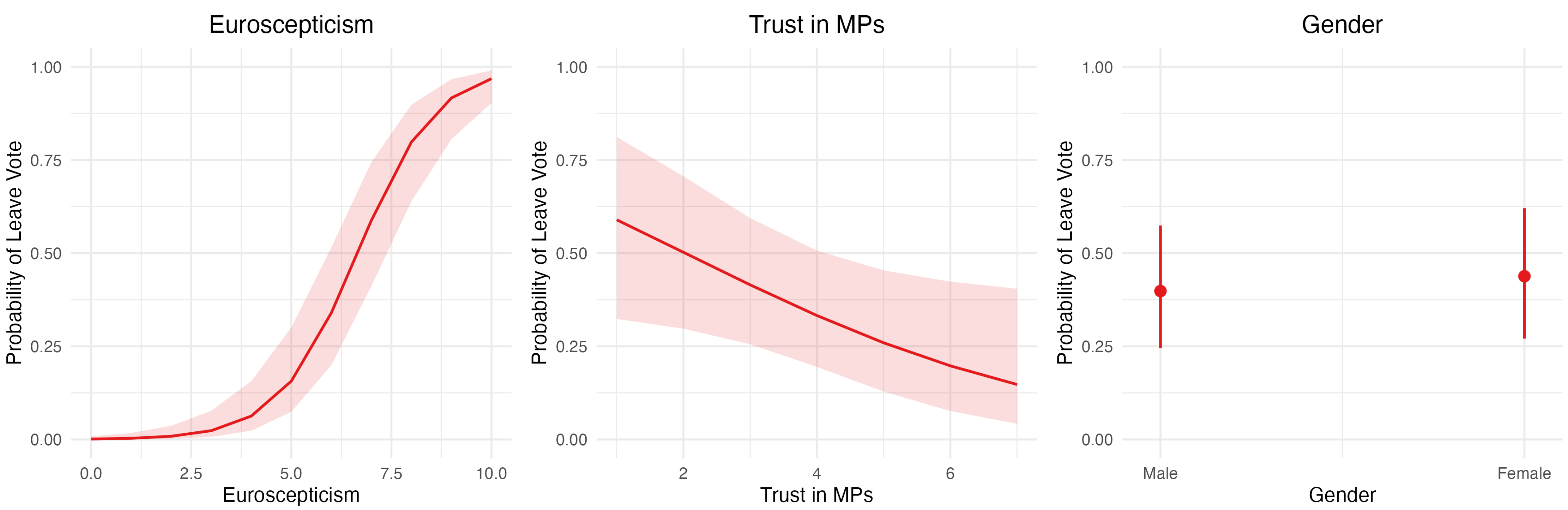
* Better (**but not best**) approach: predicted probabilities.
* Hold all other variables constant to their mean (if numeric) or to their reference category (if nominal); get the predicted probability and its confidence intervals. Use `ggpredict` from the `ggeffects` package to get the values, or `plot_model` function from `sjPlot` to get the plot.

Visualising Predicted Probabilities

- Better (**but not best**) approach: predicted probabilities.
- Hold all other variables constant to their mean (if numeric) or to their reference category (if nominal); get the predicted probability and its confidence intervals. Use `ggpredict` from the `ggeffects` package to get the values, or `plot_model` function from `sjPlot` to get the plot.
- $\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1 \text{Euroscepticism} + \beta_2 \text{Trust} + \beta_3 \text{Gender})$

Visualising Predicted Probabilities

- Better (**but not best**) approach: predicted probabilities.
- Hold all other variables constant to their mean (if numeric) or to their reference category (if nominal); get the predicted probability and its confidence intervals. Use `ggpredict` from the `ggeffects` package to get the values, or `plot_model` function from `sjPlot` to get the plot.
- $\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1 \text{Euroscepticism} + \beta_2 \text{Trust} + \beta_3 \text{Gender})$



Logistic Regression in R

```
> model <- glm(Leave ~ Euroscepticism + likeJohnson, data = bes,
+     family = "binomial")
>
> model <- glm(Leave ~ Euroscepticism + likeJohnson, data = bes,
+     family = binomial(link = "logit"))
>
> model <- glm(Euroscepticism ~  likeJohnson, data = bes,
+              family = "binomial")
Error in eval(family$initialize) : y values must be 0 <= y <= 1
>
```

Logistic Regression: Minimal Aims

Logistic Regression: Minimal Aims

- Use when the dependent variable is a $0-1$ binary variable, and we want to know the **probability** that it takes the value of 1.

Logistic Regression: Minimal Aims

* Use when the dependent variable is a $0 - 1$ binary variable, and we want to know the **probability** that it takes the value of 1.
* It fits a curve **constrained between 0 and 1**. You can do with your independent variables all the things we've seen with OLS (add multiple controls, categorical predictors, interactions, polynomials).

Logistic Regression: Minimal Aims

- Use when the dependent variable is a $0 - 1$ binary variable, and we want to know the **probability** that it takes the value of 1.
- It fits a curve **constrained between 0 and 1**. You can do with your independent variables all the things we've seen with OLS (add multiple controls, categorical predictors, interactions, polynomials).
- Coefficients refer to predicted change **in log-odds**: not easily interpretable.

Logistic Regression: Minimal Aims

* Use when the dependent variable is a $0 - 1$ binary variable, and we want to know the **probability** that it takes the value of 1.
* It fits a curve **constrained between 0 and 1**. You can do with your independent variables all the things we've seen with OLS (add multiple controls, categorical predictors, interactions, polynomials).
* Coefficients refer to predicted change **in log-odds**: not easily interpretable.
* But you can safely interpret **sign** ($\pm$) and **significance**: "Euroscepticism is **positively** and **significantly** ($p < 0.05$) associated with probability of Leave vote, holding all else constant."

Logistic Regression: Minimal Aims

* Use when the dependent variable is a $0 - 1$ binary variable, and we want to know the **probability** that it takes the value of 1.
* It fits a curve **constrained between 0 and 1**. You can do with your independent variables all the things we've seen with OLS (add multiple controls, categorical predictors, interactions, polynomials).
* Coefficients refer to predicted change **in log-odds**: not easily interpretable.
* But you can safely interpret **sign** ($\pm$) and **significance**: "Euroscepticism is **positively** and **significantly** ($p < 0.05$) associated with probability of Leave vote, holding all else constant."
* Use **predicted values plot** to get a sense of substantive effects for an 'average' observation, expressed in terms of predicted probabilities.

Logistic Regression: Two Extra Steps

Logistic Regression: Two Extra Steps

- **Average Marginal Effects**
  - Summarise average relationship between the regressors and the outcome in terms of probability.
  - Useful quantity to **interpret** of model estimates, and increasingly common. But not integral or specific to logistic regression.

Logistic Regression: Two Extra Steps

* **Average Marginal Effects**
  * Summarise average relationship between the regressors and the outcome in terms of probability.
  * Useful quantity to **interpret** of model estimates, and increasingly common. But not integral or specific to logistic regression.
* **Maximum Likelihood Estimation**
  * How your statistical software picks a particular set of coefficients (i.e. a particular 'squiggle') over all possible others.
  * Essential to the **computation** of model estimates. But R does it for you, so it's just nice to have a vague idea of what's going on.

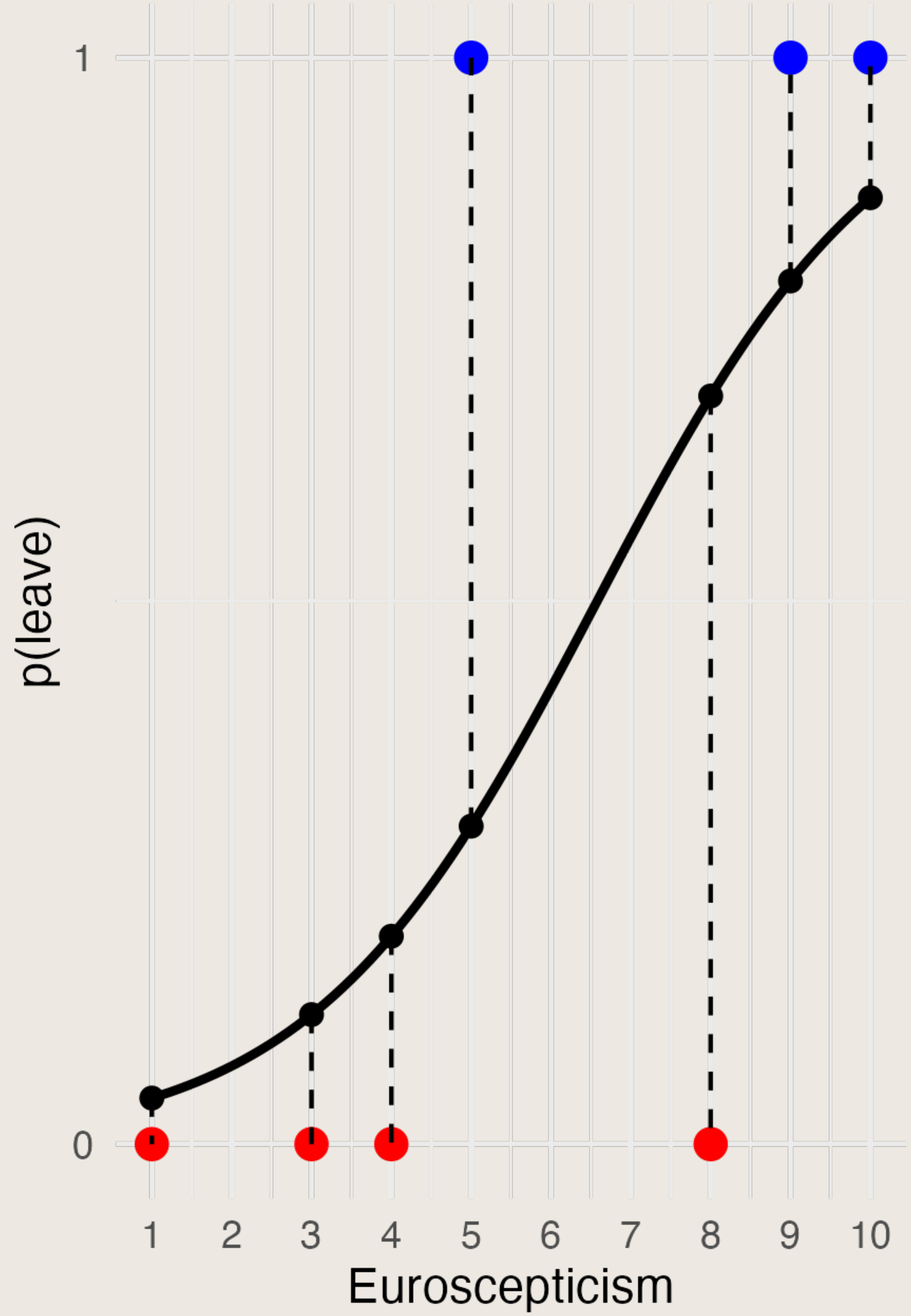
Average Marginal Effects

Average Marginal Effects

- For an observation *i*, we can get the **marginal effect**, or the **instantaneous rate of change** in probability at one point with the derivative at its predicted value.

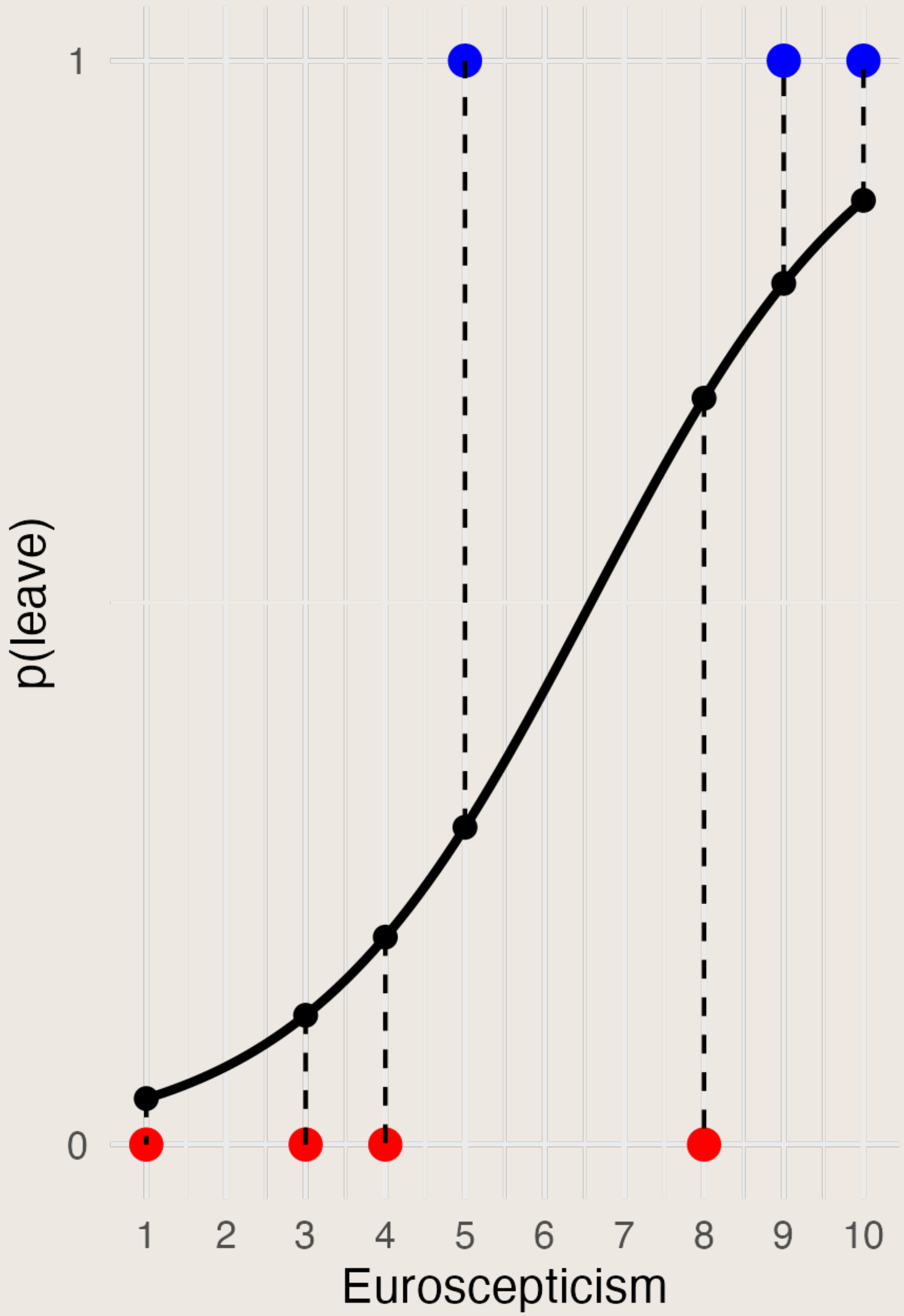
Average Marginal Effects

* For an observation *i*, we can get the **marginal effect**, or the **instantaneous rate of change** in probability at one point with the derivative at its predicted value.



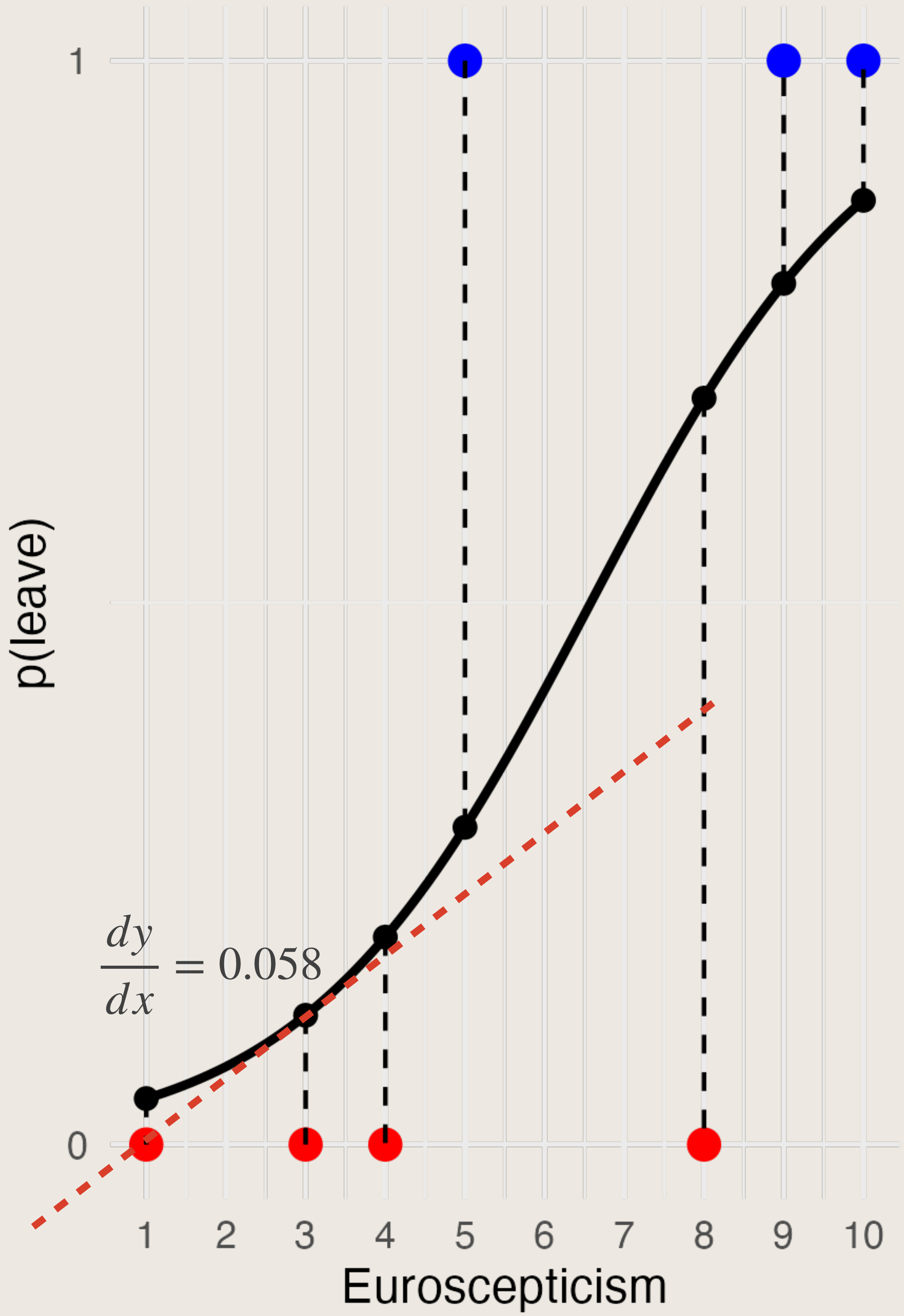
Average Marginal Effects

- For an observation *i*, we can get the **marginal effect**, or the **instantaneous rate of change** in probability at one point with the derivative at its predicted value.
- Graphically, this corresponds to the **slope of the tangent** to the curve at the predicted/fitted value $\hat{p}_i$.



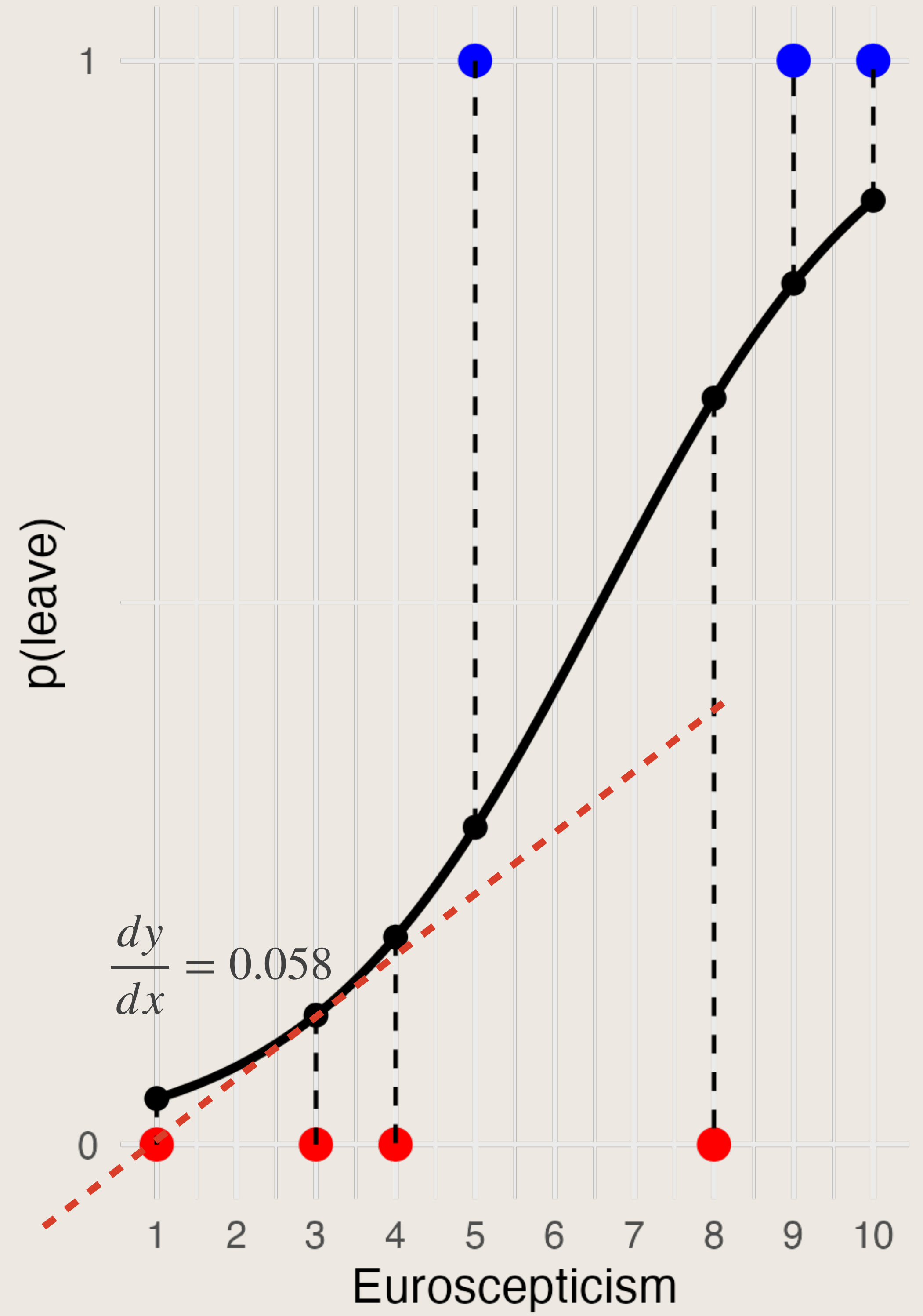
Average Marginal Effects

- For an observation *i*, we can get the **marginal effect**, or the **instantaneous rate of change** in probability at one point with the derivative at its predicted value.
- Graphically, this corresponds to the **slope of the tangent** to the curve at the predicted/fitted value $\hat{p}_i$.



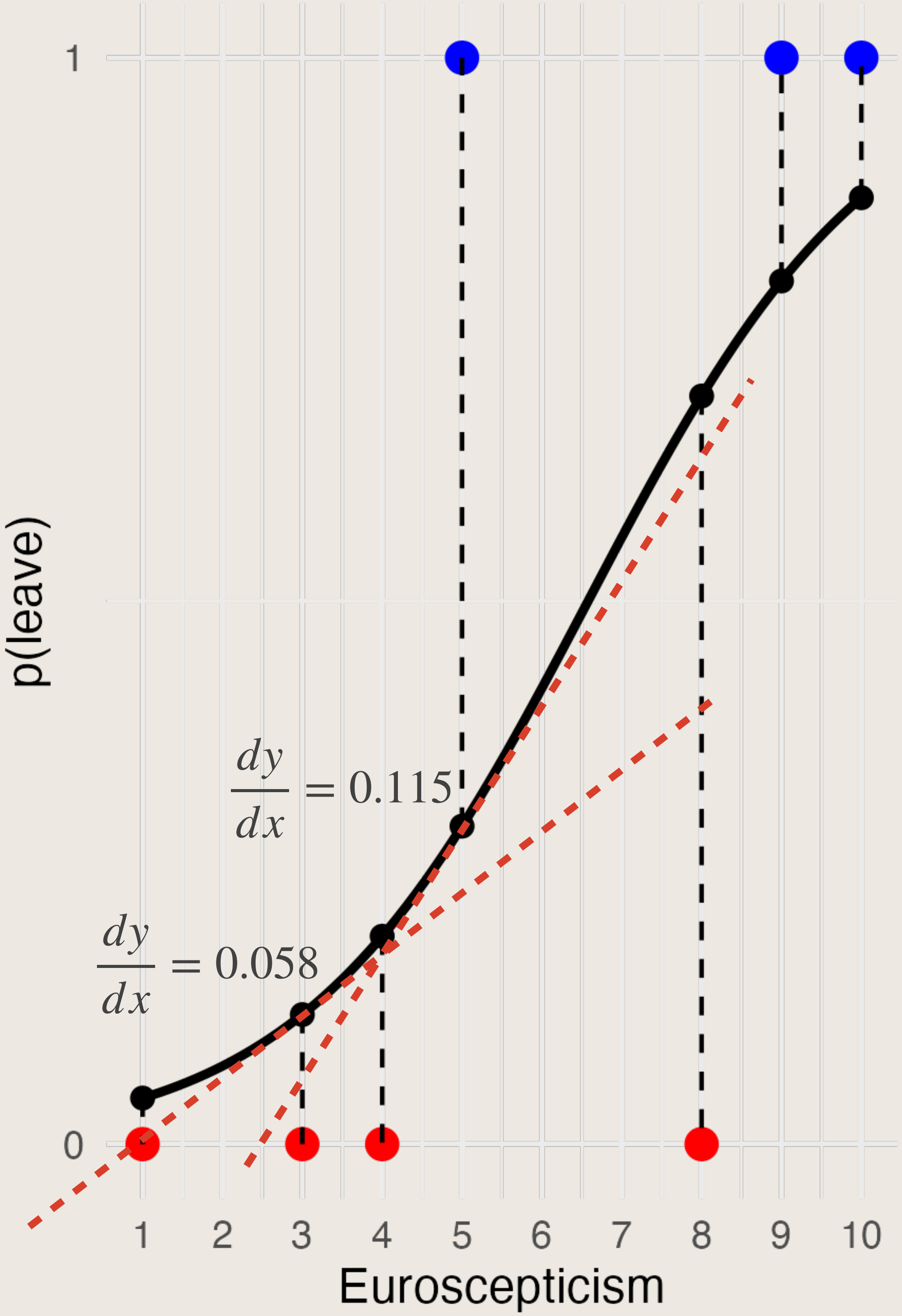
Average Marginal Effects

- For an observation *i*, we can get the **marginal effect**, or the **instantaneous rate of change** in probability at one point with the derivative at its predicted value.
- Graphically, this corresponds to the **slope of the tangent** to the curve at the predicted/fitted value $\hat{p}_i$.
- The **average marginal effects** are the average of the slopes for **all** observations in the sample.



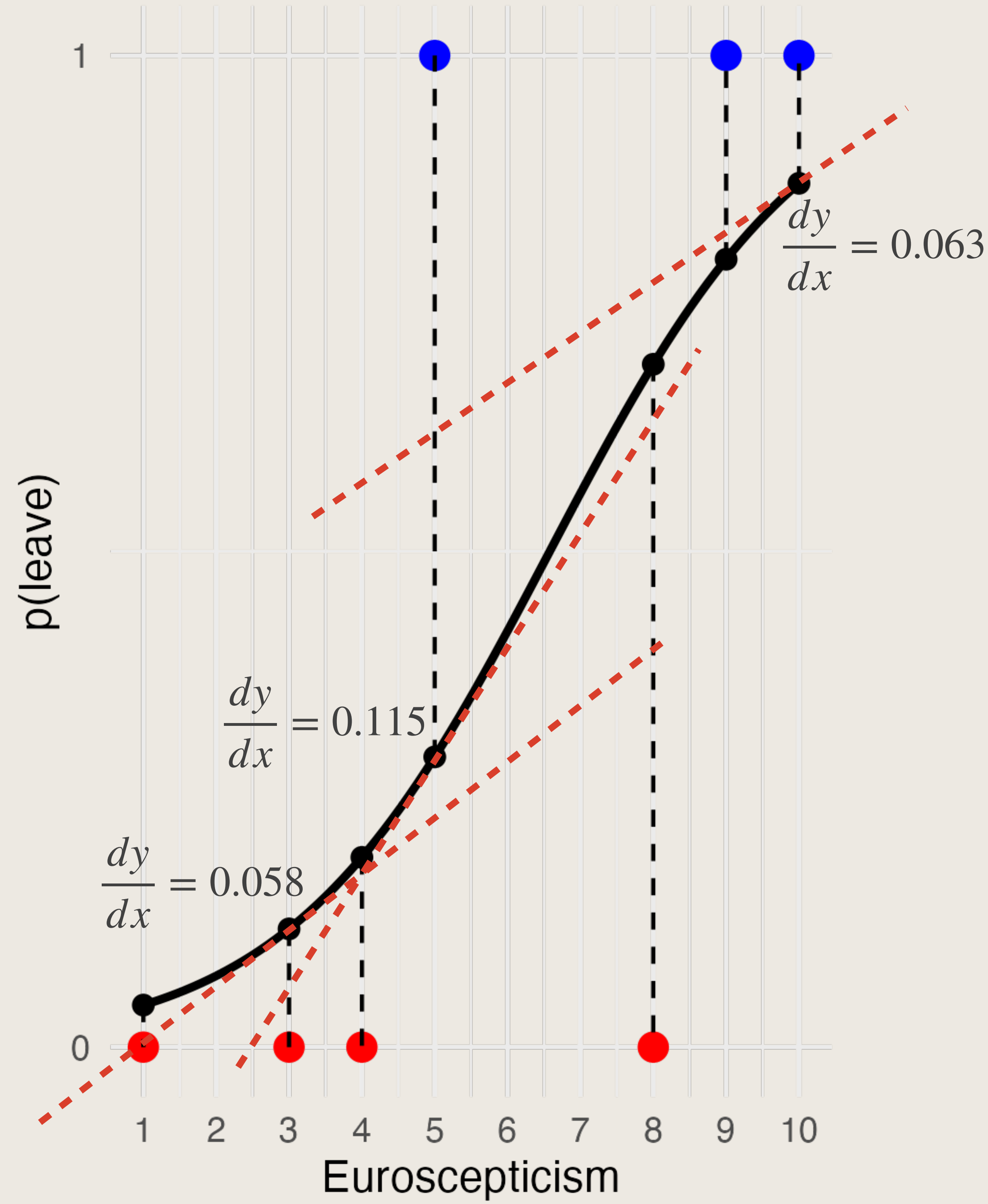
Average Marginal Effects

- For an observation *i*, we can get the **marginal effect**, or the **instantaneous rate of change** in probability at one point with the derivative at its predicted value.
- Graphically, this corresponds to the **slope of the tangent** to the curve at the predicted/fitted value $\hat{p}_i$.
- The **average marginal effects** are the average of the slopes for **all** observations in the sample.



Average Marginal Effects

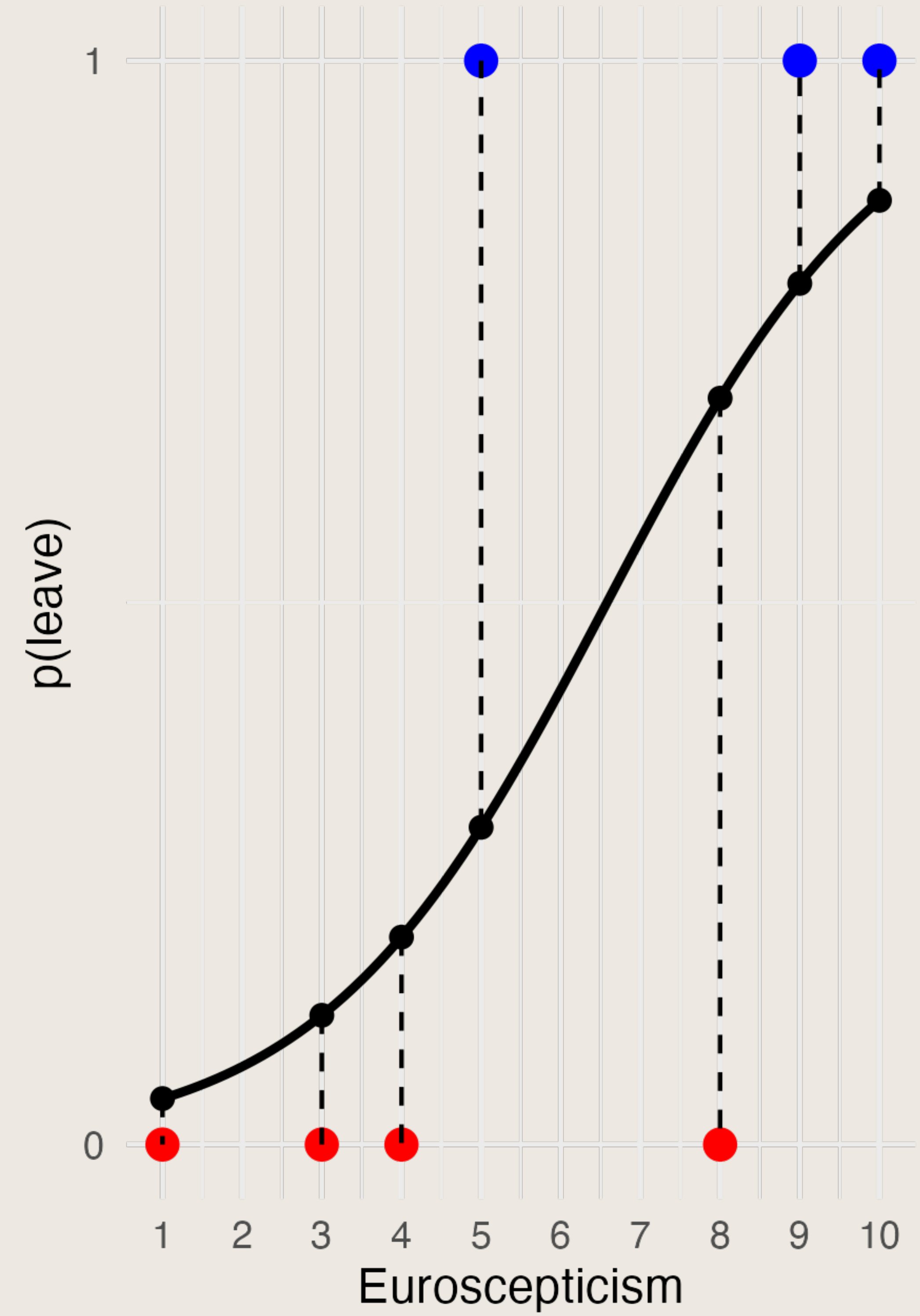
- For an observation *i*, we can get the **marginal effect**, or the **instantaneous rate of change** in probability at one point with the derivative at its predicted value.
- Graphically, this corresponds to the **slope of the tangent** to the curve at the predicted/fitted value $\hat{p}_i$.
- The **average marginal effects** are the average of the slopes for **all** observations in the sample.



Average Marginal Effects

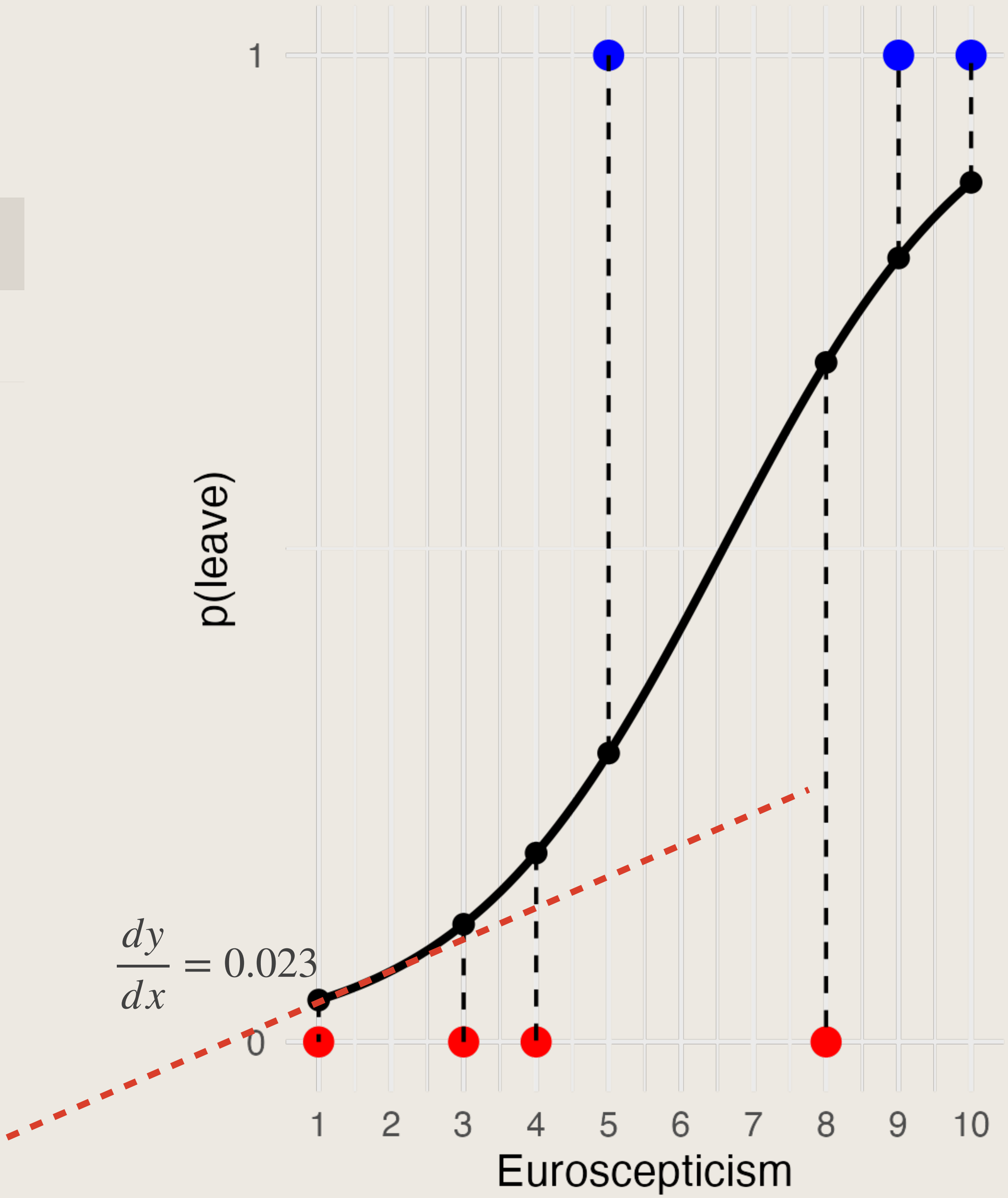
Average Marginal Effects

X	Marginal Effects



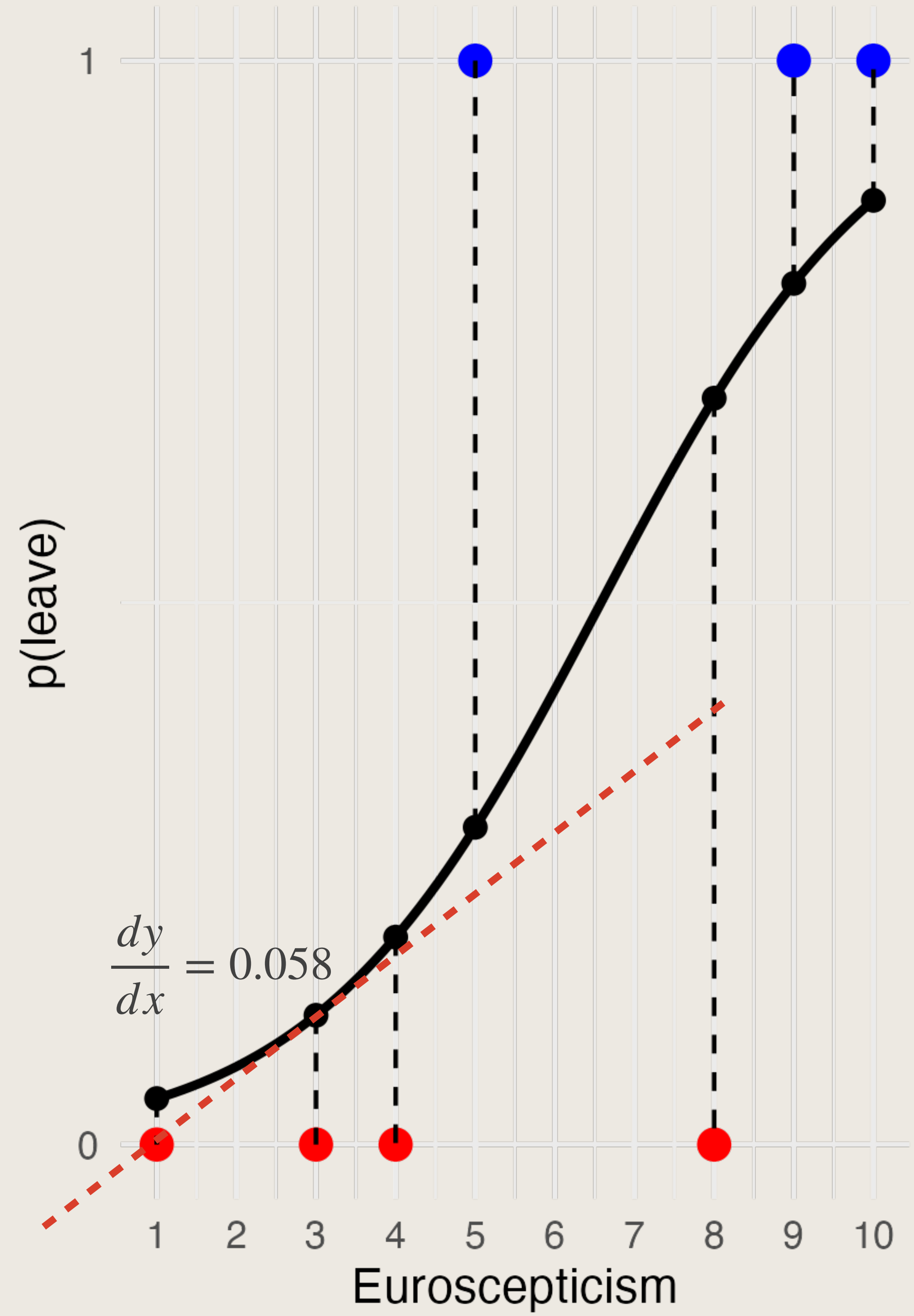
Average Marginal Effects

X	Marginal Effects
1	0.023



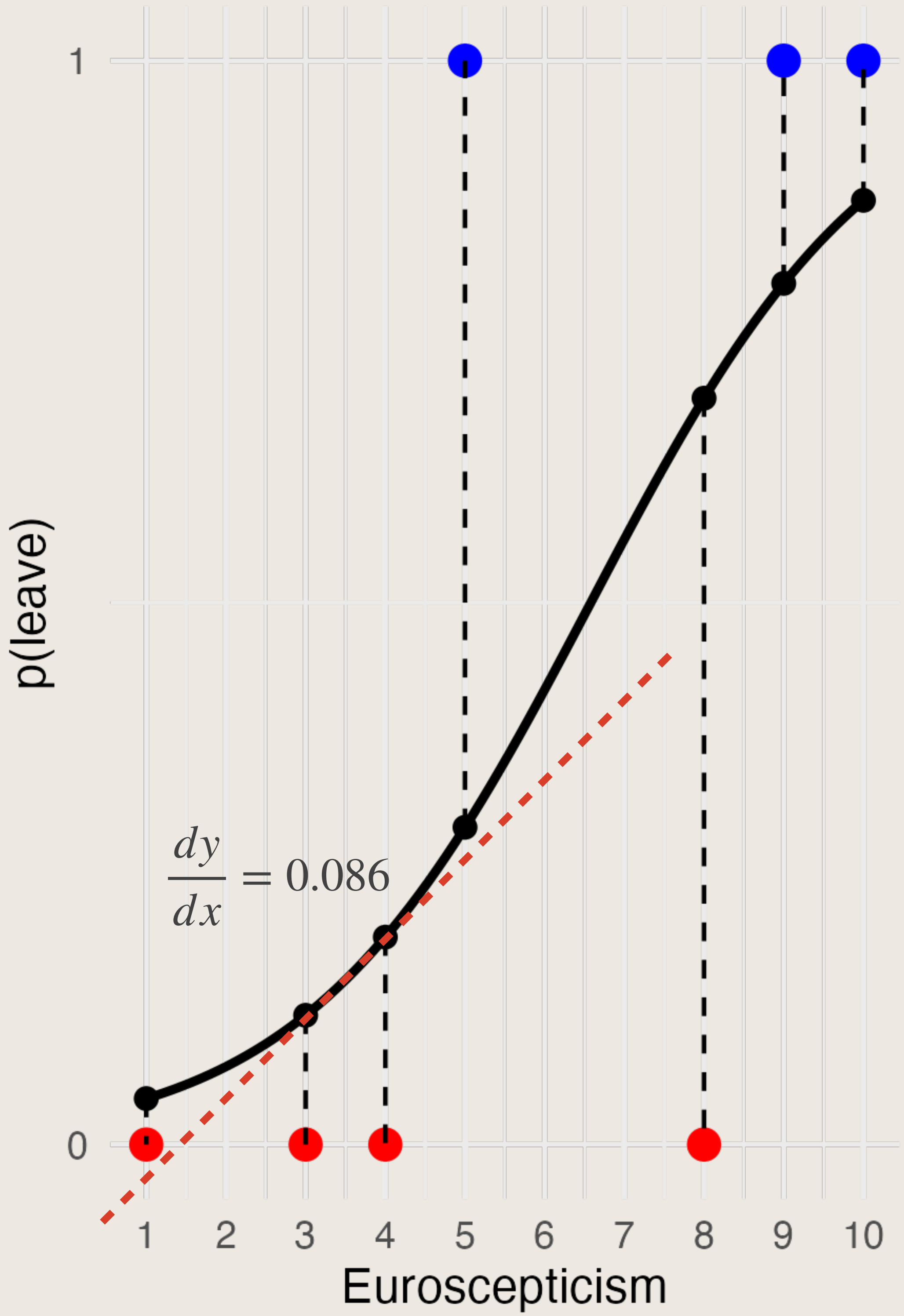
Average Marginal Effects

X	Marginal Effects
1	0.023
3	0.058



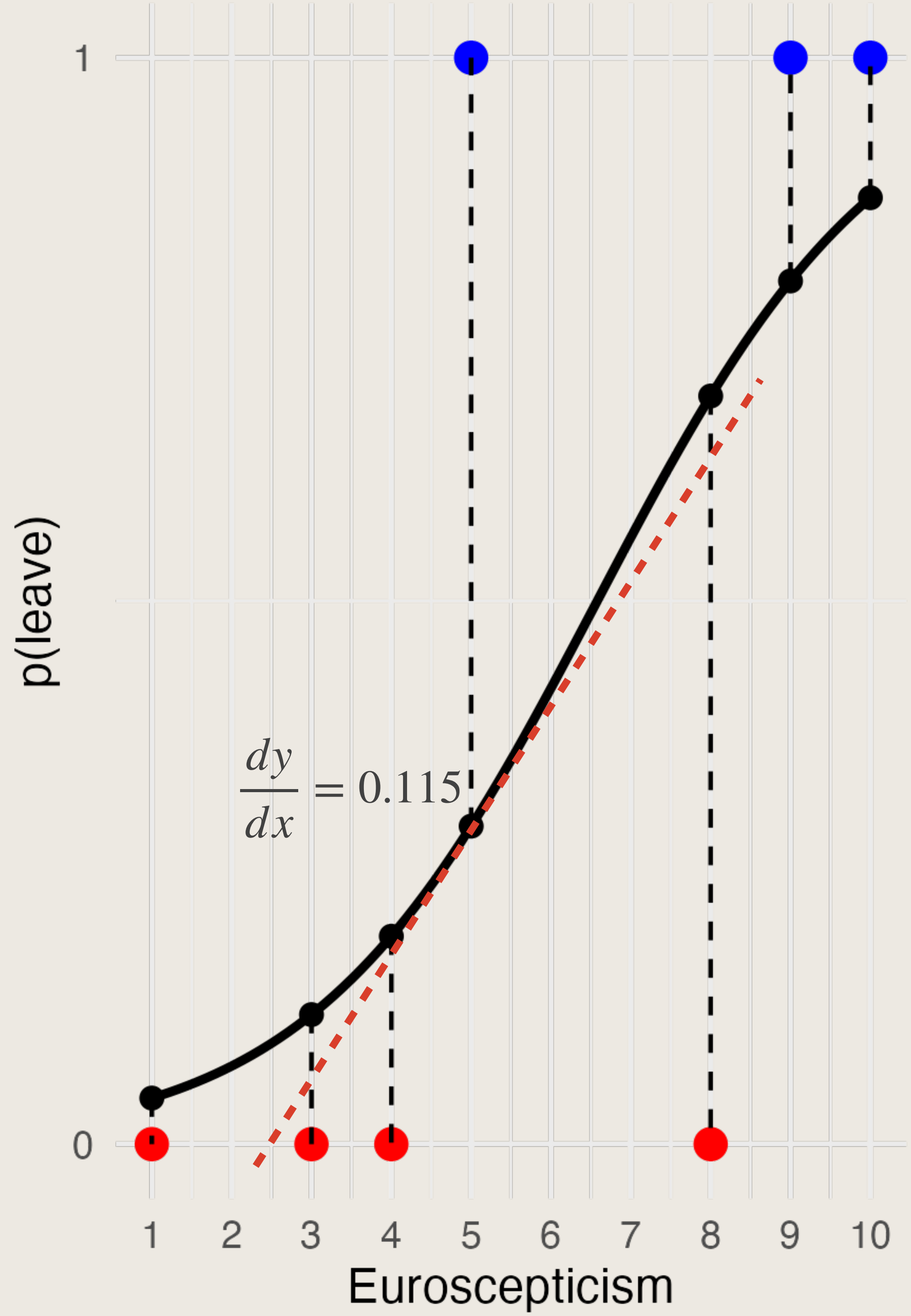
Average Marginal Effects

X	Marginal Effects
1	0.023
3	0.058
4	0.086



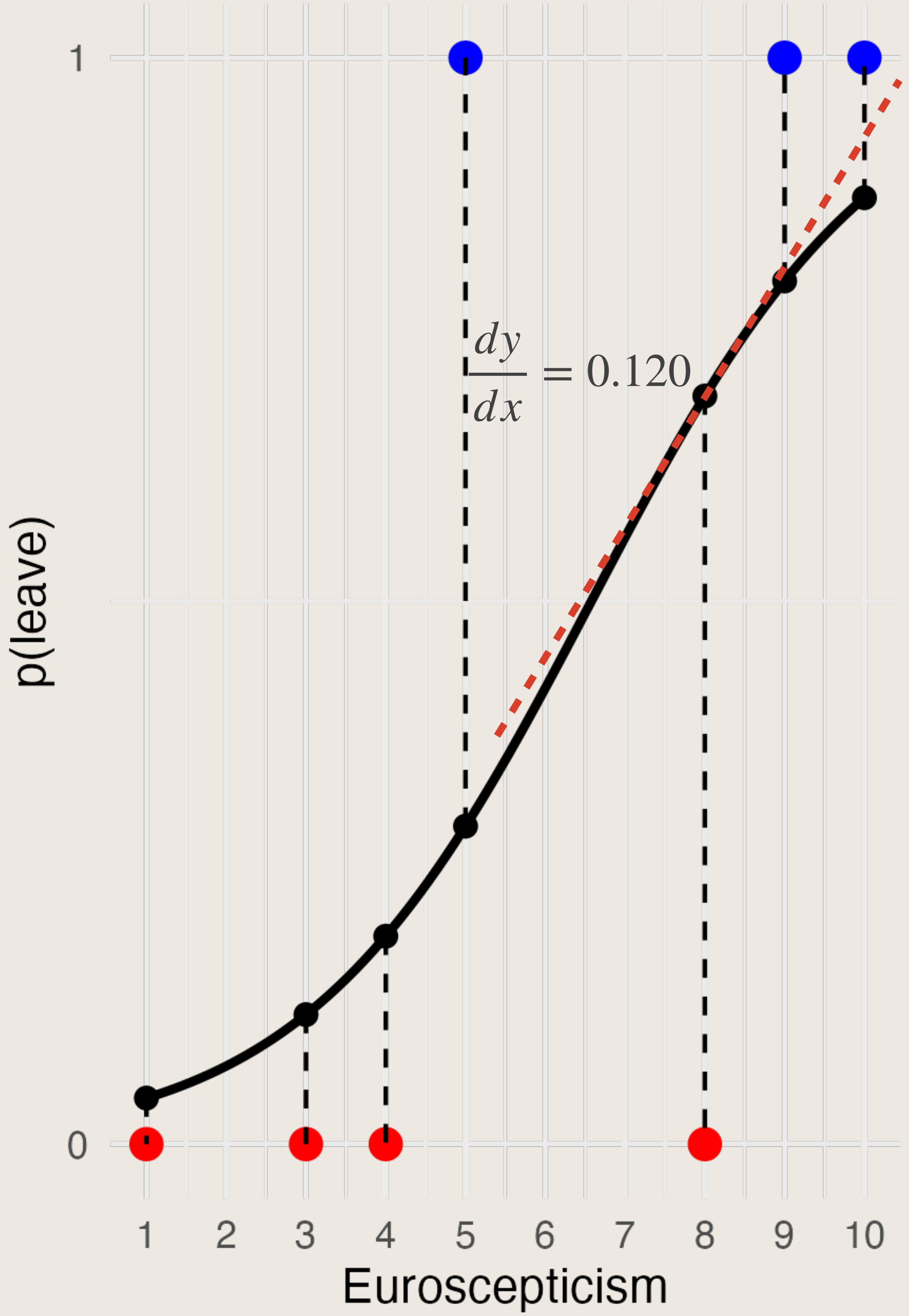
Average Marginal Effects

X	Marginal Effects
1	0.023
3	0.058
4	0.086
5	0.115



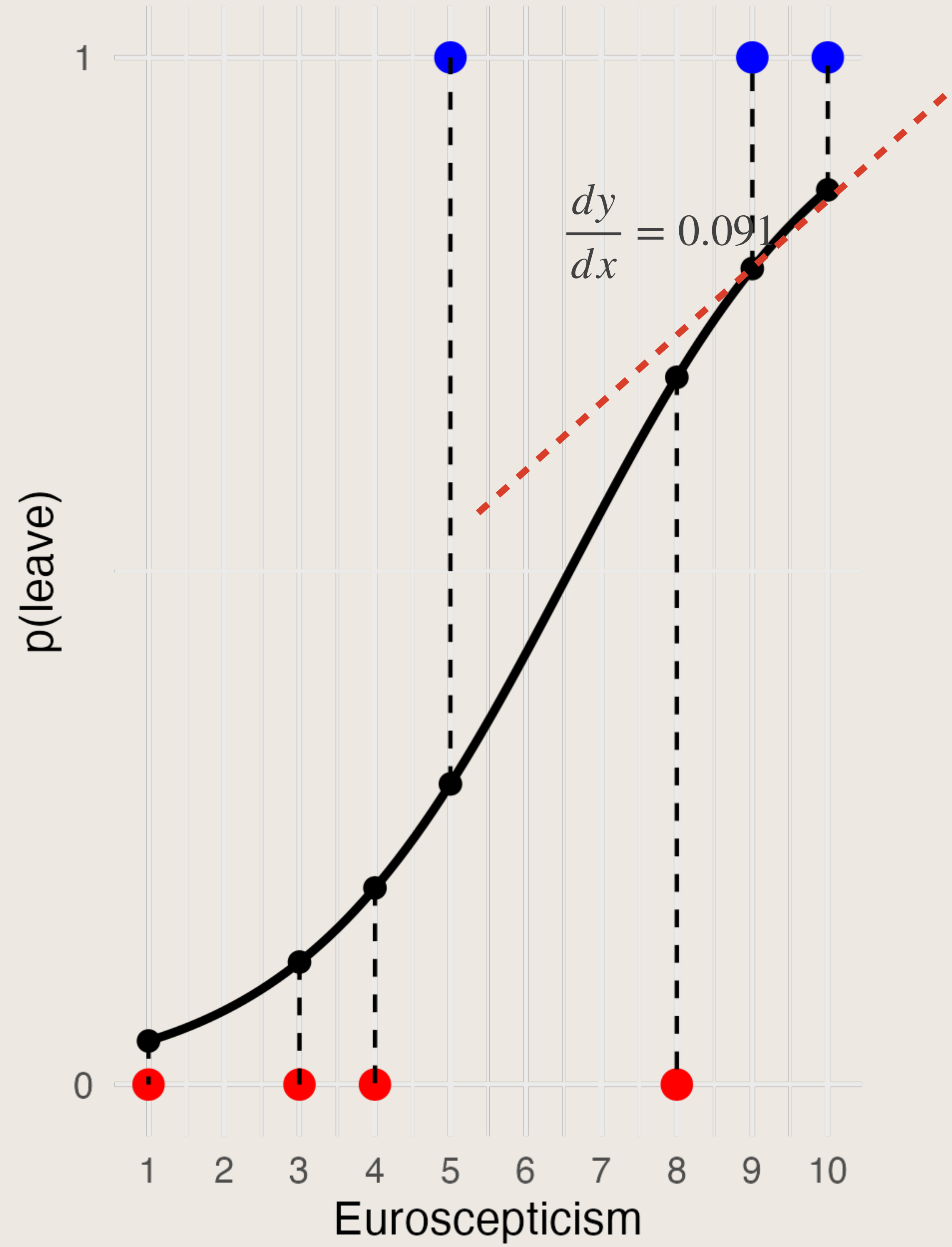
Average Marginal Effects

X	Marginal Effects
1	0.023
3	0.058
4	0.086
5	0.115
8	0.120



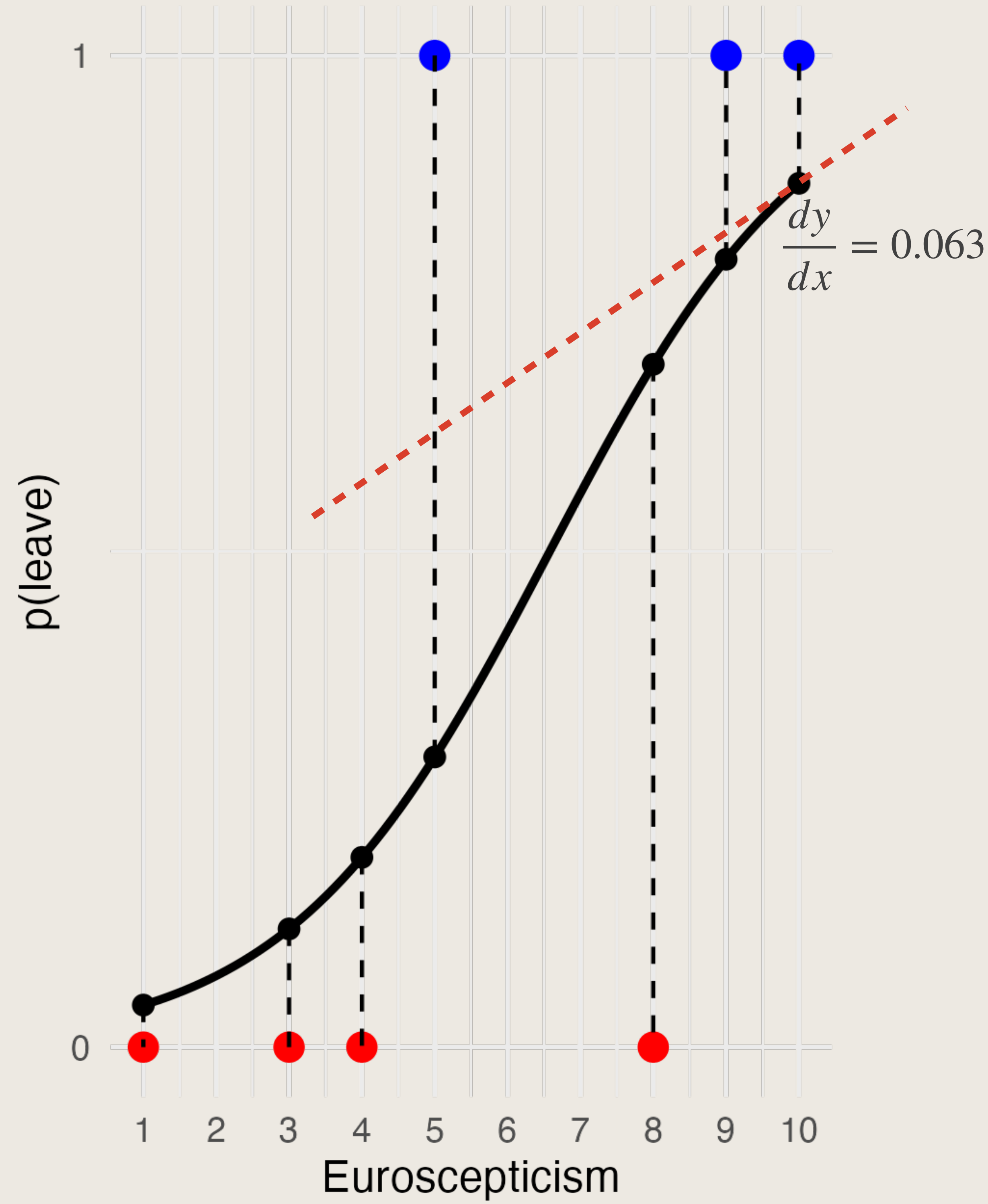
Average Marginal Effects

X	Marginal Effects
1	0.023
3	0.058
4	0.086
5	0.115
8	0.120
9	0.091



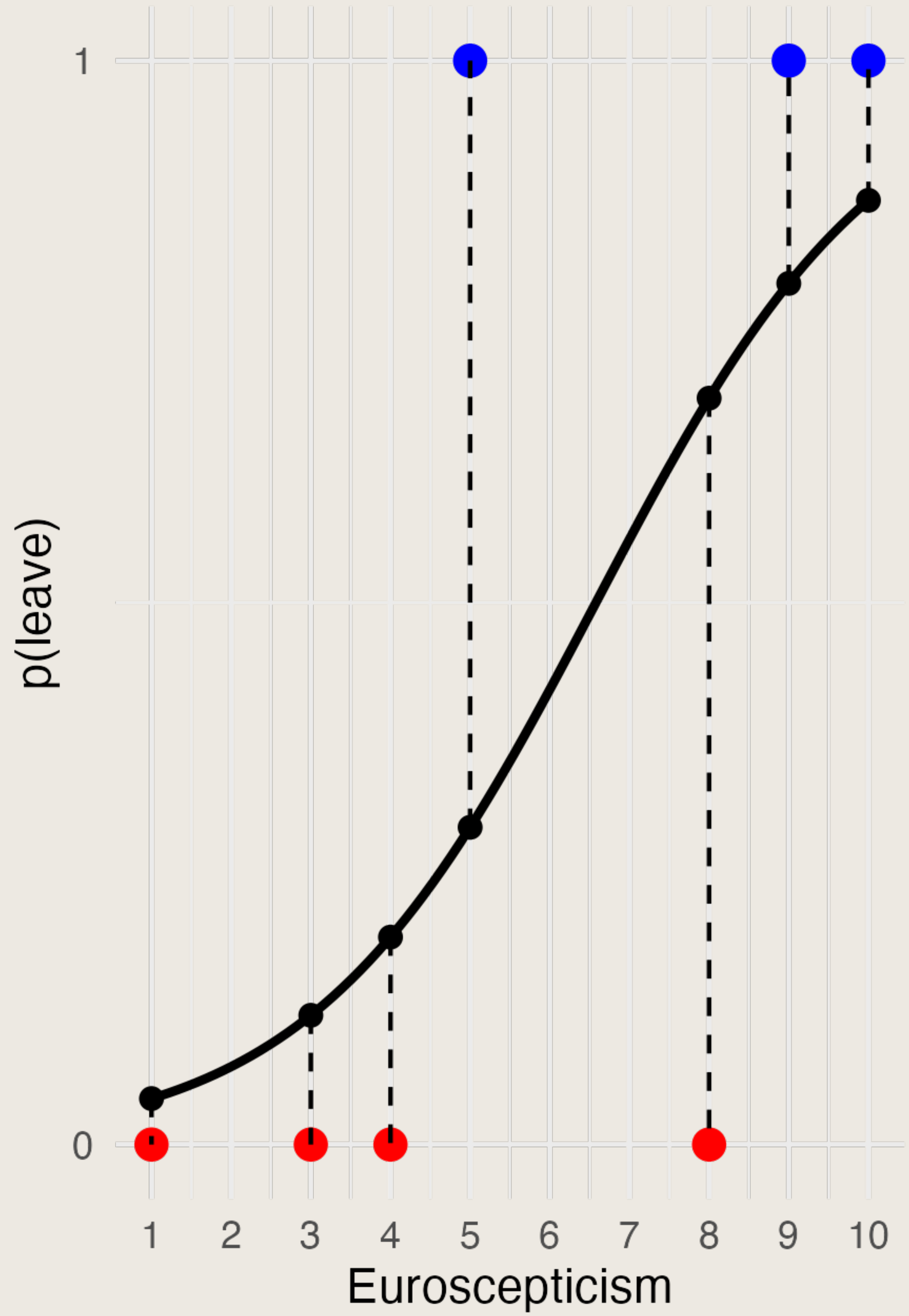
Average Marginal Effects

X	Marginal Effects
1	0.023
3	0.058
4	0.086
5	0.115
8	0.120
9	0.091
10	0.063



Average Marginal Effects

X	Marginal Effects
1	0.023
3	0.058
4	0.086
5	0.115
8	0.120
9	0.091
10	0.063
Mean	0.080



Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1\text{Euroscepticism} + \beta_2\text{Trust} + \beta_3\text{Gender} + \epsilon)$$

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1 \text{Euroscepticism} + \beta_2 \text{Trust} + \beta_3 \text{Gender} + \epsilon)$$

R#	Euroscepticism	Trust in MPs	Female	M.E. (Eurosc.)	M.E. (Trust in MPs)	ME (Female)

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1\text{Euroscepticism} + \beta_2\text{Trust} + \beta_3\text{Gender} + \epsilon)$$

R#	**Euroscepticism**	**Trust in MPs**	**Female**	**M.E. (Eurosc.)**	**M.E. (Trust in MPs)**	**ME (Female)**
1	3	2	0	0.035	—0.12	0.006

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1\text{Euroscepticism} + \beta_2\text{Trust} + \beta_3\text{Gender} + \epsilon)$$

R#	**Euroscepticism**	**Trust in MPs**	**Female**	**M.E. (Eurosc.)**	**M.E. (Trust in MPs)**	**ME (Female)**
1	3	2	0	0.035	—0.12	0.006
2	10	2	0	0.021	—0.007	0.003

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1 \text{Euroscepticism} + \beta_2 \text{Trust} + \beta_3 \text{Gender} + \epsilon)$$

R#	**Euroscepticism**	**Trust in MPs**	**Female**	**M.E. (Eurosc.)**	**M.E. (Trust in MPs)**	**ME (Female)**
1	3	2	0	0.035	−0.12	0.006
2	10	2	0	0.021	−0.007	0.003
3	2	4	1	0.008	−0.003	0.001

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1\text{Euroscepticism} + \beta_2\text{Trust} + \beta_3\text{Gender} + \epsilon)$$

R#	**Euroscepticism**	**Trust in MPs**	**Female**	**M.E. (Willingness)**	**M.E. (Trust in MPs)**	**ME (Female)**
1	3	2	0	0.035	—0.12	0.006
2	10	2	0	0.021	—0.007	0.003
3	2	4	1	0.008	—0.003	0.001
4	10	5	1	0.049	—0.017	0.008

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1 \text{Euroscepticism} + \beta_2 \text{Trust} + \beta_3 \text{Gender} + \epsilon)$$

R#	Euroscepticism	Trust in MPs	Female	M.E. (Eurosc.)	M.E. (Trust in MPs)	ME (Female)
1	3	2	0	0.035	—0.12	0.006
2	10	2	0	0.021	—0.007	0.003
3	2	4	1	0.008	—0.003	0.001
4	10	5	1	0.049	—0.017	0.008
5	10	4	0	0.041	—0.014	0.006

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1\text{Euroscepticism} + \beta_2\text{Trust} + \beta_3\text{Gender} + \epsilon)$$

R#	**Euroscepticism**	**Trust in MPs**	**Female**	**M.E. (Eurosc.)**	**M.E. (Trust in MPs)**	**ME (Female)**
1	3	2	0	0.035	—0.12	0.006
2	10	2	0	0.021	—0.007	0.003
3	2	4	1	0.008	—0.003	0.001
4	10	5	1	0.049	—0.017	0.008
5	10	4	0	0.041	—0.014	0.006
6	0	3	1	0.001	—0.001	0.0002

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1 \text{Euroscepticism} + \beta_2 \text{Trust} + \beta_3 \text{Gender} + \epsilon)$$

R#	**Euroscepticism**	**Trust in MPs**	**Female**	**M.E. (Eurosc.)**	**M.E. (Trust in MPs)**	**ME (Female)**
1	3	2	0	0.035	—0.12	0.006
2	10	2	0	0.021	—0.007	0.003
3	2	4	1	0.008	—0.003	0.001
4	10	5	1	0.049	—0.017	0.008
5	10	4	0	0.041	—0.014	0.006
6	0	3	1	0.001	—0.001	0.0002
7	7	4	1	0.251	—0.087	0.041

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1\text{Euroscepticism} + \beta_2\text{Trust} + \beta_3\text{Gender} + \epsilon)$$

R#	**Euroscepticism**	**Trust in MPs**	**Female**	**M.E. (Eurosc.)**	**M.E. (Trust in MPs)**	**ME (Female)**
1	3	2	0	0.035	−0.12	0.006
2	10	2	0	0.021	−0.007	0.003
3	2	4	1	0.008	−0.003	0.001
4	10	5	1	0.049	−0.017	0.008
5	10	4	0	0.041	−0.014	0.006
6	0	3	1	0.001	−0.001	0.0002
7	7	4	1	0.251	−0.087	0.041
8	10	4	0	0.041	−0.014	0.006

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1 \text{Euroscepticism} + \beta_2 \text{Trust} + \beta_3 \text{Gender} + \epsilon)$$

R#	**Euroscepticism**	**Trust in MPs**	**Female**	**M.E. (Eurosc.)**	**M.E. (Trust in MPs)**	**ME (Female)**
1	3	2	0	0.035	—0.12	0.006
2	10	2	0	0.021	—0.007	0.003
3	2	4	1	0.008	—0.003	0.001
4	10	5	1	0.049	—0.017	0.008
5	10	4	0	0.041	—0.014	0.006
6	0	3	1	0.001	—0.001	0.0002
7	7	4	1	0.251	—0.087	0.041
8	10	4	0	0.041	—0.014	0.006
…	…	…	…	…	…	…

Average Marginal Effects

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1\text{Euroscepticism} + \beta_2\text{Trust} + \beta_3\text{Gender} + \epsilon)$$

R#	Euroscepticism	Trust in MPs	Female	M.E. (Eurosc.)	M.E. (Trust in MPs)	ME (Female)
1	3	2	0	0.035	—0.12	0.006
2	10	2	0	0.021	—0.007	0.003
3	2	4	1	0.008	—0.003	0.001
4	10	5	1	0.049	—0.017	0.008
5	10	4	0	0.041	—0.014	0.006
6	0	3	1	0.001	—0.001	0.0002
7	7	4	1	0.251	—0.087	0.041
8	10	4	0	0.041	—0.014	0.006
…	…	…	…	…	…	…
Average Marginal Effects (=Mean)				0.078	—0.027	0.013

Average Marginal Effects in R

$$\Pr(\text{Leave}) = \text{logit}^{-1}(\alpha + \beta_1\text{Euroscepticism} + \beta_2\text{Trust} + \beta_3\text{Gender} + \epsilon)$$

```
> margins(model)
Average marginal effects
glm(formula = Leave ~ Euroscepticism + trustMPs + gender, family = "binomial",     data = bes)

 Euroscepticism trustMPs  gender
        0.07841 -0.02719 0.01259
> margins_summary(model)
         factor     AME     SE       z      p   lower   upper
 Euroscepticism  0.0784 0.0014 55.9855 0.0000  0.0757  0.0812
         gender  0.0126 0.0394  0.3195 0.7493 -0.0646  0.0898
       trustMPs -0.0272 0.0128 -2.1195 0.0340 -0.0523 -0.0020
> head(marginal_effects(model))
  dydx_Euroscepticism dydx_trustMPs  dydx_gender
1         0.034934821 -0.0140799805 0.0056085222
2         0.021075526 -0.0086488138 0.0033835164
3         0.007761427 -0.0032417654 0.0012460383
4         0.048529412 -0.0192219974 0.0077910323
5         0.040877210 -0.0163496473 0.0065625284
6         0.001454886 -0.0006128382 0.0002335708
> 
```

Average Marginal Effects

Average Marginal Effects

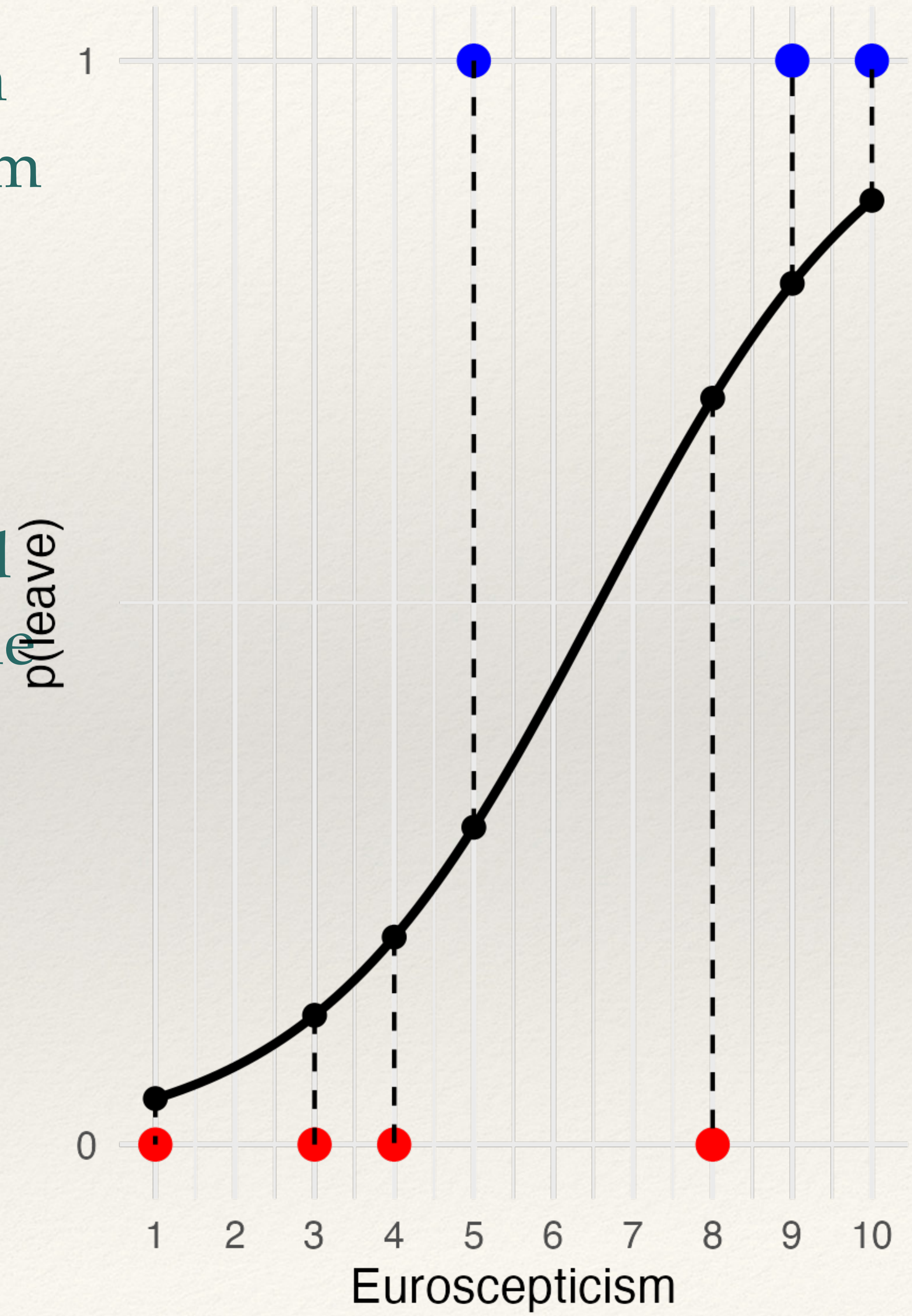
- **Almost-correct** interpretation of AMEs: “on average, a 1-point increase in Euroscepticism is associated with a 7.8 percentage-point increase in probability of voting Leave.”

Average Marginal Effects

- **Almost-correct** interpretation of AMEs: "on average, a 1-point increase in Euroscepticism is associated with a 7.8 percentage-point increase in probability of voting Leave."
- Why '*almost*' correct? Because AMEs aren't averages of one-point changes (these would be slopes that go from $\hat{Y}_i$ for X_i and the value of $\hat{Y}_i$ for $X_i + 1$).

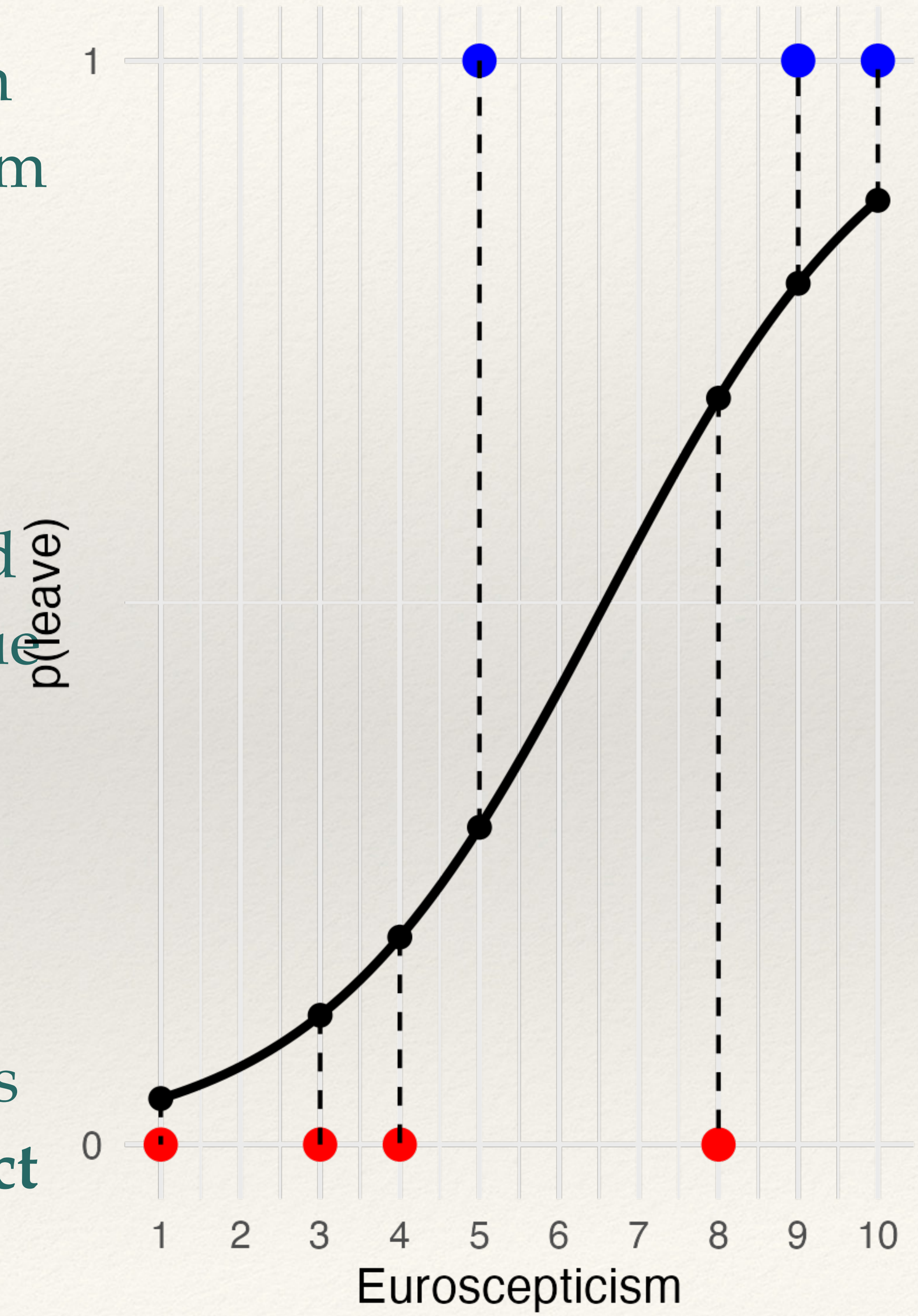
Average Marginal Effects

- **Almost-correct** interpretation of AMEs: "on average, a 1-point increase in Euroscepticism is associated with a 7.8 percentage-point increase in probability of voting Leave."
- Why '*almost*' correct? Because AMEs aren't averages of one-point changes (these would be slopes that go from $\hat{Y}_i$ for X_i and the value of $\hat{Y}_i$ for $X_i + 1$).
- AMEs are averages of slopes **at one point** (i.e. derivatives).



Average Marginal Effects

- **Almost-correct** interpretation of AMEs: "on average, a 1-point increase in Euroscepticism is associated with a 7.8 percentage-point increase in probability of voting Leave."
- Why '*almost*' correct? Because AMEs aren't averages of one-point changes (these would be slopes that go from $\hat{Y}_i$ for X_i and the value of $\hat{Y}_i$ for $X_i + 1$).
- AMEs are averages of slopes **at one point** (i.e. derivatives).
- But the 'one-point increase' interpretation is fine. Or just say: the **average marginal effect** of Euroscepticism is 7.8 percentage points.



Average Marginal Effects

Average Marginal Effects

* Common these days to have a marginal effects plot alongside of a regression table of (largely unintelligible) log-odds coefficients.

Average Marginal Effects

* Common these days to have a marginal effects plot alongside of a regression table of (largely unintelligible) log-odds coefficients.
* AMEs estimates are **superior** to alternative approaches (included the predicted probabilities we plotted earlier), because they are calculated on data from our sample, not from hypothetical ‘average cases’.

Average Marginal Effects

* Common these days to have a marginal effects plot alongside of a regression table of (largely unintelligible) log-odds coefficients.
* AMEs estimates are **superior** to alternative approaches (included the predicted probabilities we plotted earlier), because they are calculated on data from our sample, not from hypothetical 'average cases'.
  * See: Hanmer, M.J. and Kalkan, K. (2013) "Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models." *AJPS*, 57(1), pp.263-277.

ffects

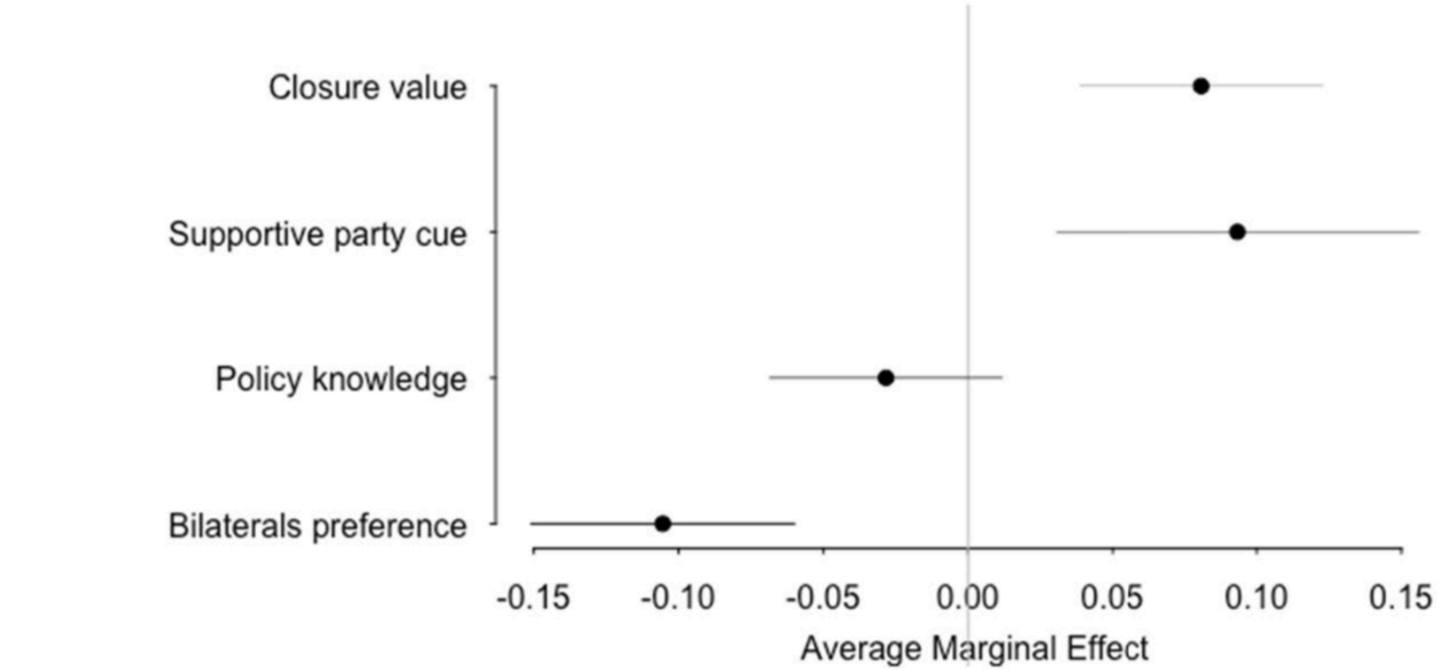


Fig. 2 Average marginal effects on IMI-support
Note: plot displays the average marginal effects with 95% confidence intervals for the independent variables on IMI support. Estimates based on a logistic regression model with standardized independent variables (for detailed model output see Model 1 in Table A-2 in the supplementary materials)

Armingeon and Lutz (2022)

: alongside of a regression
s.

ches (included the
they are calculated on
data from our sample, not from hypothetical 'average cases'.

- See: Hanmer, M.J. and Kalkan, K. (2013) "Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models." *AJPS*, 57(1), pp.263-277.

...cts

...de of a regression

...luded the

... calculated on

dat... ...'.

Fig. 2 Ave...
Note: plot...
ables on I...
ables (for ...

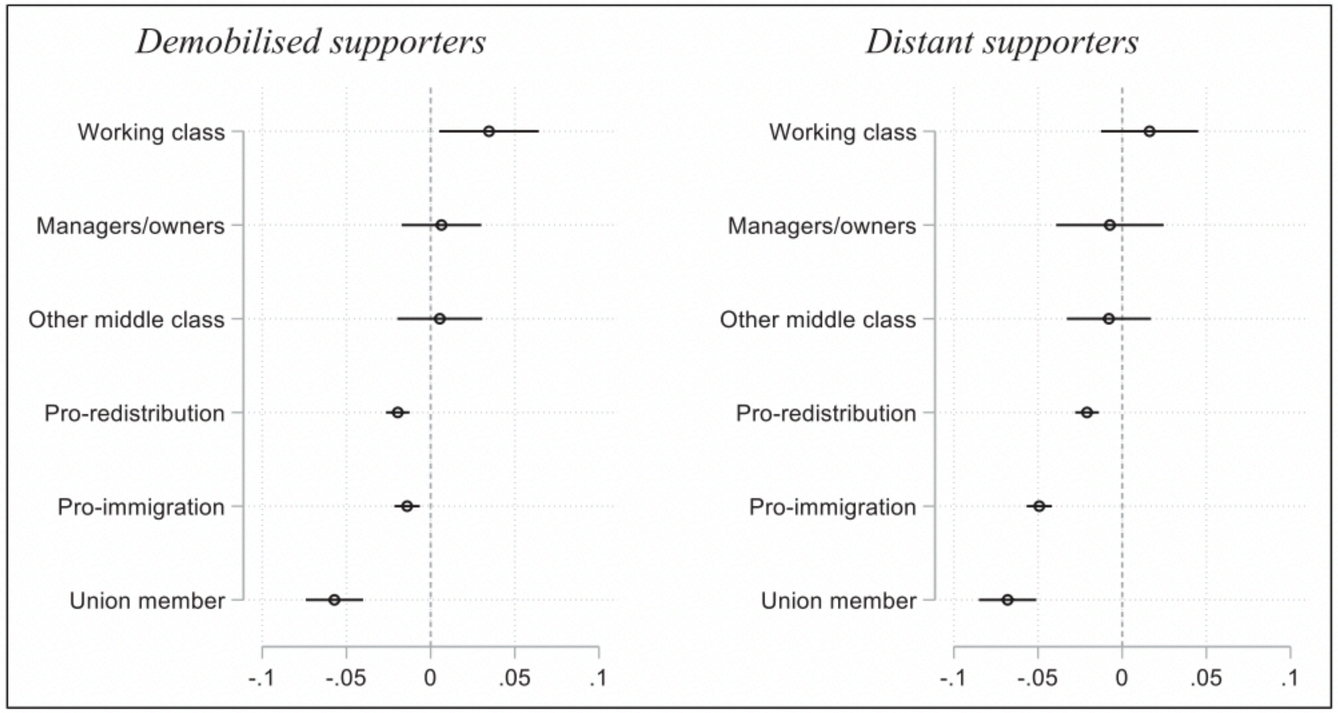
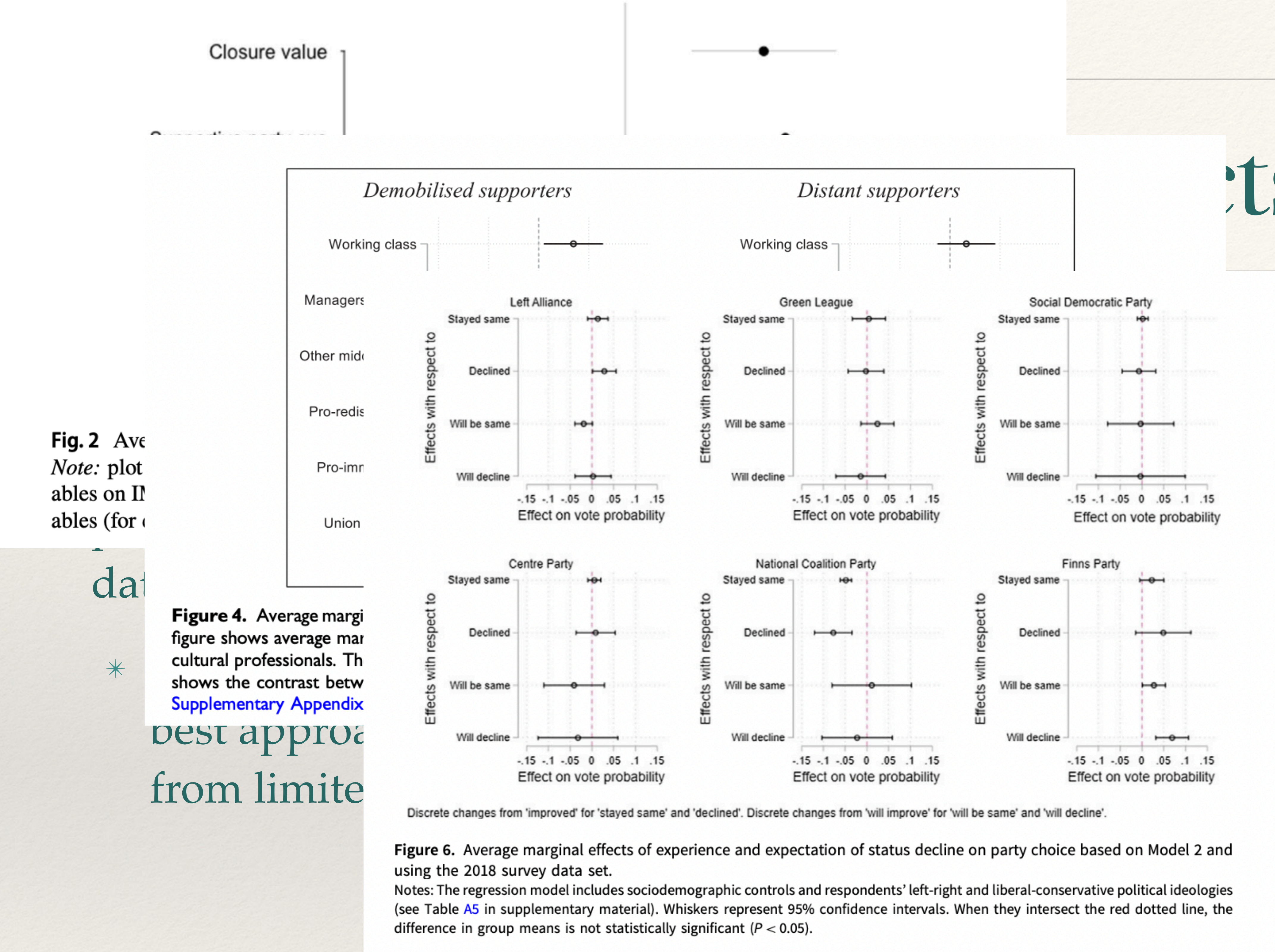


Figure 4. Average marginal effects of class, union membership, attitudes on the different support groups of social democracy. Note: The figure shows average marginal effects based on the models presented in Table B.2. The reference category for social class is socio-cultural professionals. The left-hand side shows the contrast between demobilised and core supporters, whereas the right-hand side shows the contrast between distant and core supporters. The contrast between demobilised and distant supporters is shown in Supplementary Appendix D.2.

- ...ve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models." *AJPS*, 57(1), pp.263-277.

Bremer and Rennwald (2023)

...cts



...of a regression

...led the
...lculated on
dat...

✳ ... Clarifying the
best approa... ...rginal effects
from limite... ...63-277.

Im, Wass, Kantola and Kauppinen (2022)

Fig. 2 Ave
Note: plot
ables on I
ables (for

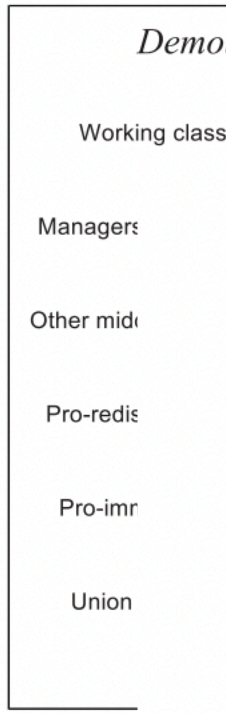


Figure 4. Average margi
figure shows average mar
cultural professionals. Th
shows the contrast betw
Supplementary Appendix

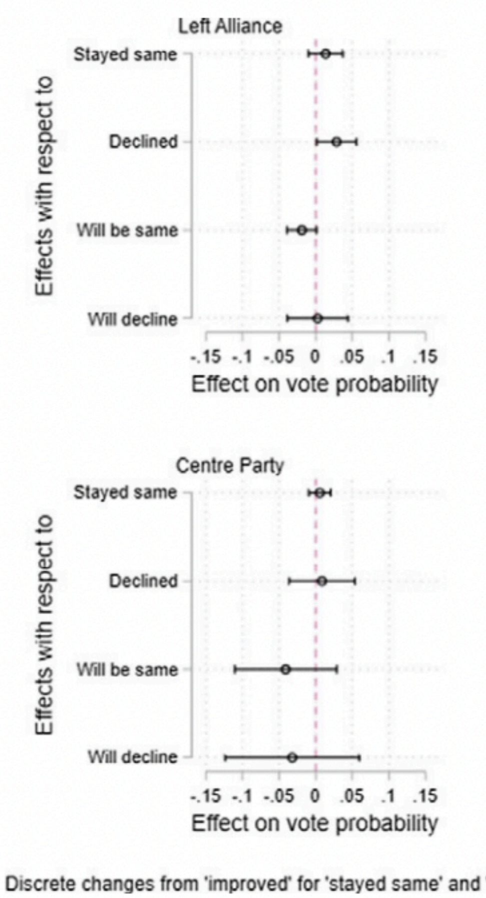


Figure 6. Average marginal effects of exper
using the 2018 survey data set.
Notes: The regression model includes sociodemo
(see Table A5 in supplementary material). Whis
difference in group means is not statistically si

dat

✳

best appro
from limite

Im, Wass, K

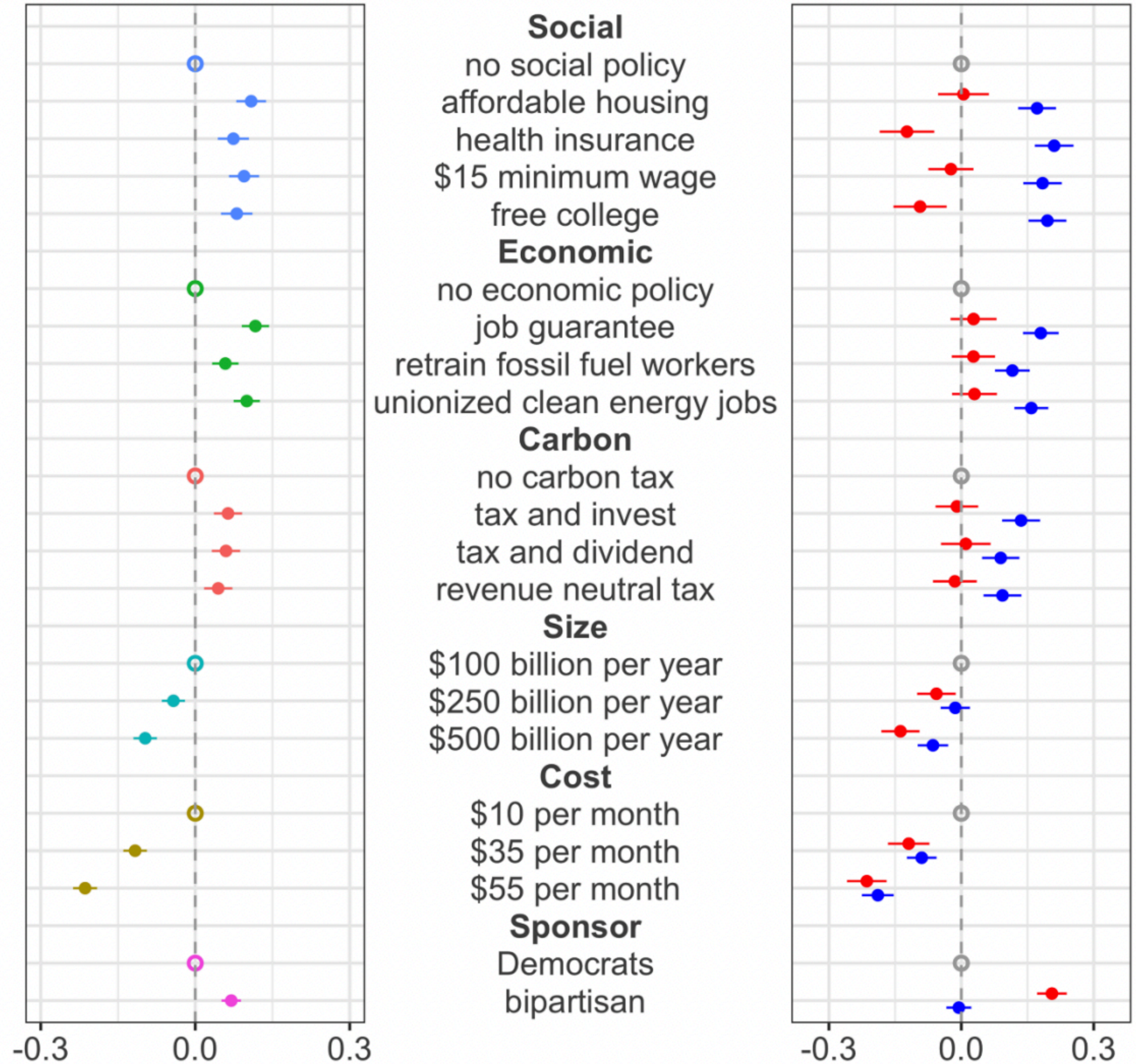


Figure 1. How social, economic, and climate programs shape support for bundled climate policy. The left panel shows average effects of each policy element (colored by policy dimension) on support for the policy bundle, while the right panel shows party-specific effects (red = Republican, blue = Democrat). Policy dimensions include carbon taxes, social programs, economic programs, energy costs, government spending levels, and party sponsorship. Point estimates are average marginal component effects (AMCEs) with 95% confidence intervals for each policy level. Each AMCE estimates how inclusion of the listed program affects support for the bundled climate package. Each element is compared against a base category for each policy dimension, denoted by an open circle.

Bergquist, Mildenberger and Stokes (2020)

Average Marginal Effects

- Common these days to have a marginal effects plot alongside of a regression table of (largely unintelligible) log-odds coefficients.
- AMEs estimates are **superior** to alternative approaches (included the predicted probabilities we plotted earlier), because they are calculated on data from our sample, not from hypothetical 'average cases'.
  - See: Hanmer, M.J. and Kalkan, K. (2013) "Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models." *AJPS*, 57(1), pp.263-277.

Average Marginal Effects

- Common these days to have a marginal effects plot alongside of a regression table of (largely unintelligible) log-odds coefficients.
- AMEs estimates are **superior** to alternative approaches (included the predicted probabilities we plotted earlier), because they are calculated on data from our sample, not from hypothetical 'average cases'.
  - See: Hanmer, M.J. and Kalkan, K. (2013) "Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models." *AJPS*, 57(1), pp.263-277.
- AMEs even more important to convey effect size estimates as your logistic models become more complex (interactions, polynomials):

Average Marginal Effects

- Common these days to have a marginal effects plot alongside of a regression table of (largely unintelligible) log-odds coefficients.
- AMEs estimates are **superior** to alternative approaches (included the predicted probabilities we plotted earlier), because they are calculated on data from our sample, not from hypothetical 'average cases'.
  - See: Hanmer, M.J. and Kalkan, K. (2013) "Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models." *AJPS*, 57(1), pp.263-277.
- AMEs even more important to convey effect size estimates as your logistic models become more complex (interactions, polynomials):
  - Mize, T.D. (2019) "Best practices for estimating, interpreting, and presenting nonlinear interaction effects." *Sociological Science*, 6, pp.81-117.

Maximum Likelihood: Motivation

Maximum Likelihood: Motivation

* So far, we've ignored the question of how we obtain the coefficients — i.e. how we picked the 'sigmoid of best fit'.

Maximum Likelihood: Motivation

- So far, we've ignored the question of how we obtain the coefficients — i.e. how we picked the 'sigmoid of best fit'.
- With OLS, we chose the line that minimises the sum of squared residuals, where the residual for observation i is defined as $\hat{\epsilon}_i = Y_i - \hat{Y}_i$.

Maximum Likelihood: Motivation

- So far, we've ignored the question of how we obtain the coefficients — i.e. how we picked the 'sigmoid of best fit'.
- With OLS, we chose the line that minimises the sum of squared residuals, where the residual for observation i is defined as $\hat{\epsilon}_i = Y_i - \hat{Y}_i$.
- We can't do the same with $\log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta \text{Euroscepticism}$:

Maximum Likelihood: Motivation

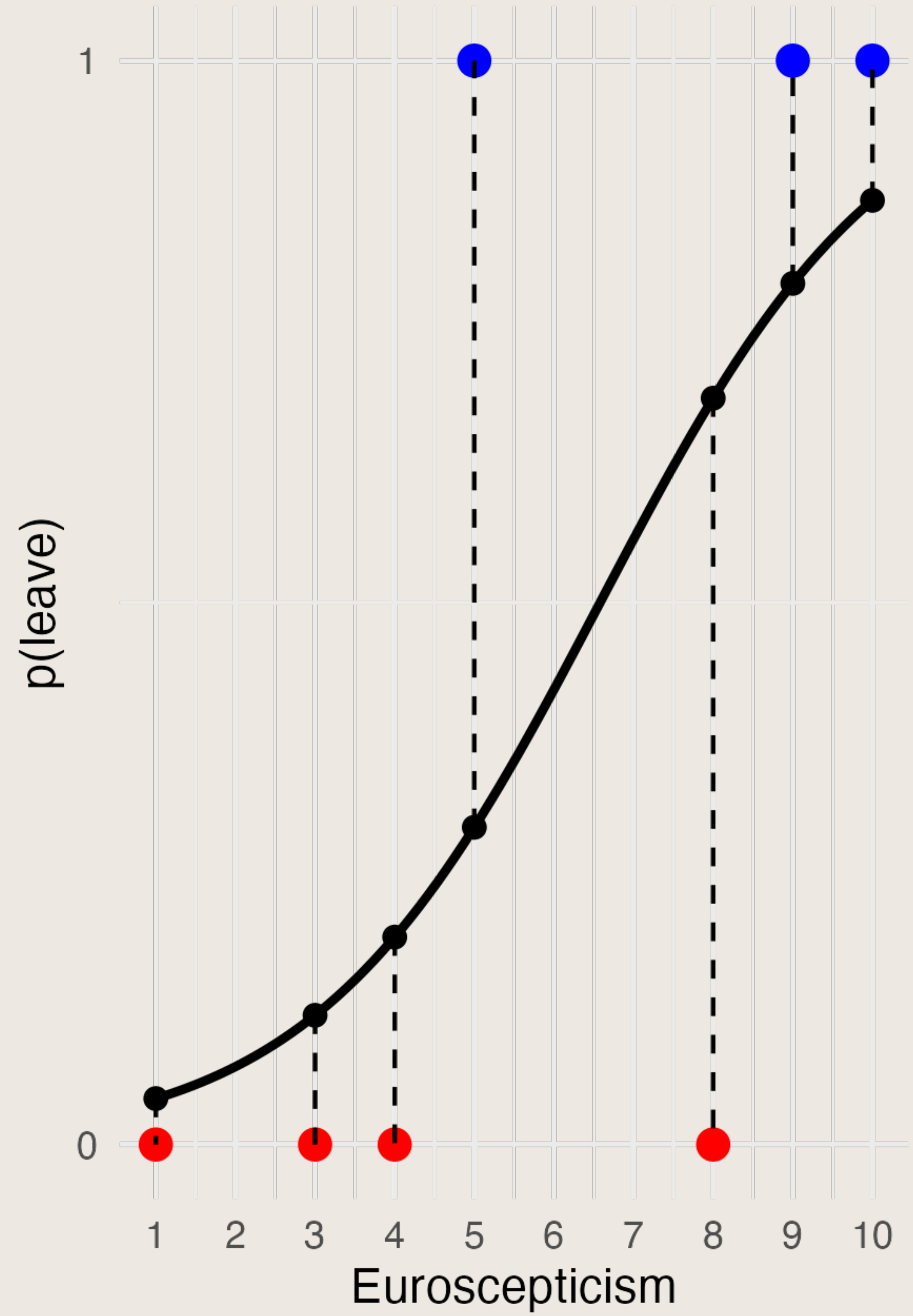
- So far, we've ignored the question of how we obtain the coefficients — i.e. how we picked the 'sigmoid of best fit'.
- With OLS, we chose the line that minimises the sum of squared residuals, where the residual for observation i is defined as $\hat{\epsilon}_i = Y_i - \hat{Y}_i$.
- We can't do the same with $\log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta \text{Euroscepticism}$:
  - When $\text{Leave}_i = 0$ we get $\log \frac{0}{1} = -\infty$ (negative infinity). So all residuals will be $-\infty - \text{something} = -\infty$.

Maximum Likelihood: Motivation

- So far, we've ignored the question of how we obtain the coefficients — i.e. how we picked the 'sigmoid of best fit'.
- With OLS, we chose the line that minimises the sum of squared residuals, where the residual for observation i is defined as $\hat{\epsilon}_i = Y_i - \hat{Y}_i$.
- We can't do the same with $\log \frac{\Pr(\text{Leave} = 1)}{\Pr(\text{Leave} = 0)} = \alpha + \beta\text{Euroscepticism}$:
  - When $\text{Leave}_i = 0$ we get $\log \frac{0}{1} = -\infty$ (negative infinity). So all residuals will be $-\infty - \text{something} = -\infty$.
  - When $\text{Leave}_i = 1$ we get $\log \frac{1}{0} = +\infty$ (positive infinity). So all residuals will be $+\infty - \text{something} = +\infty$.

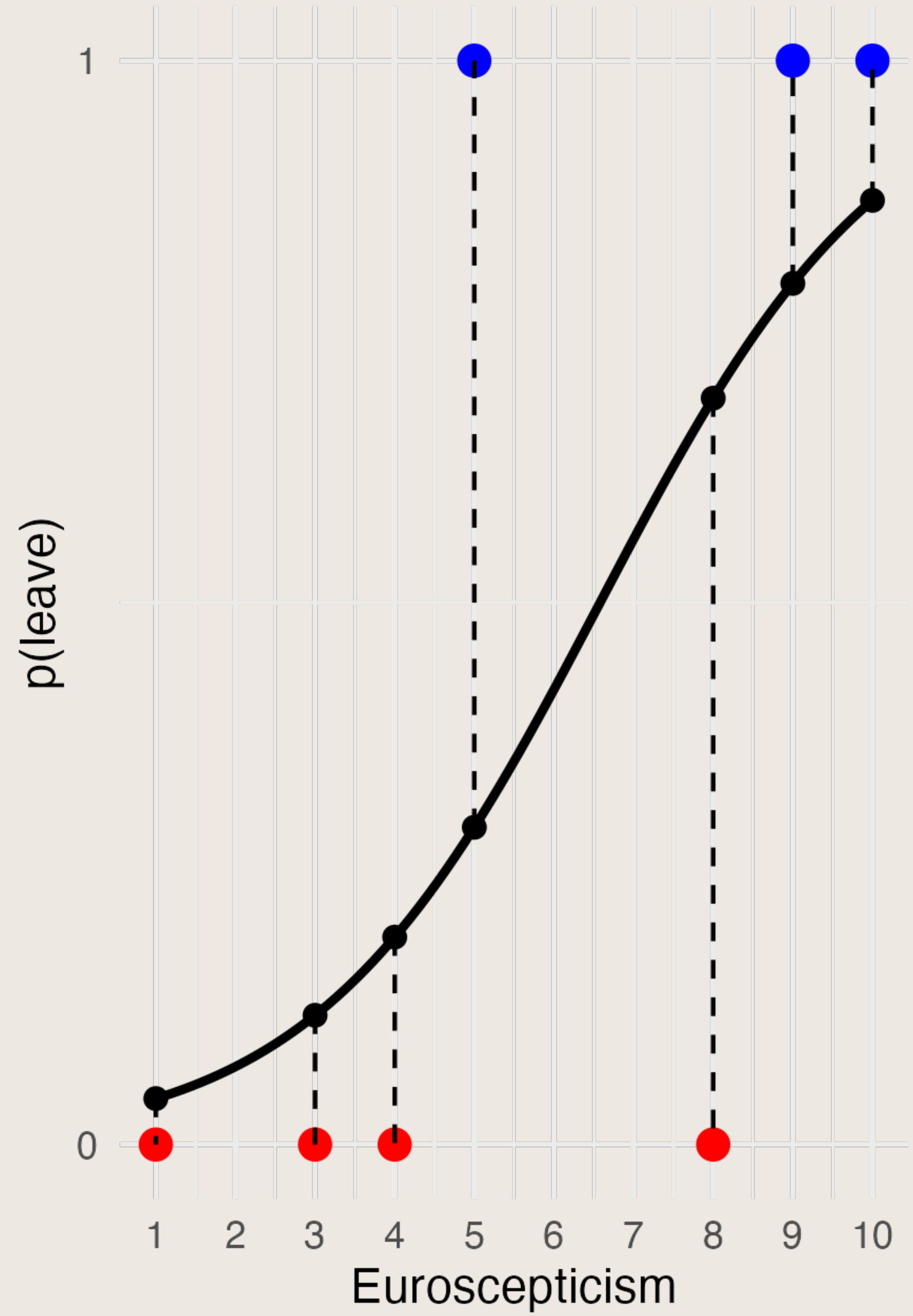
Maximum Likelihood: Intuition

Maximum Likelihood: Intuition



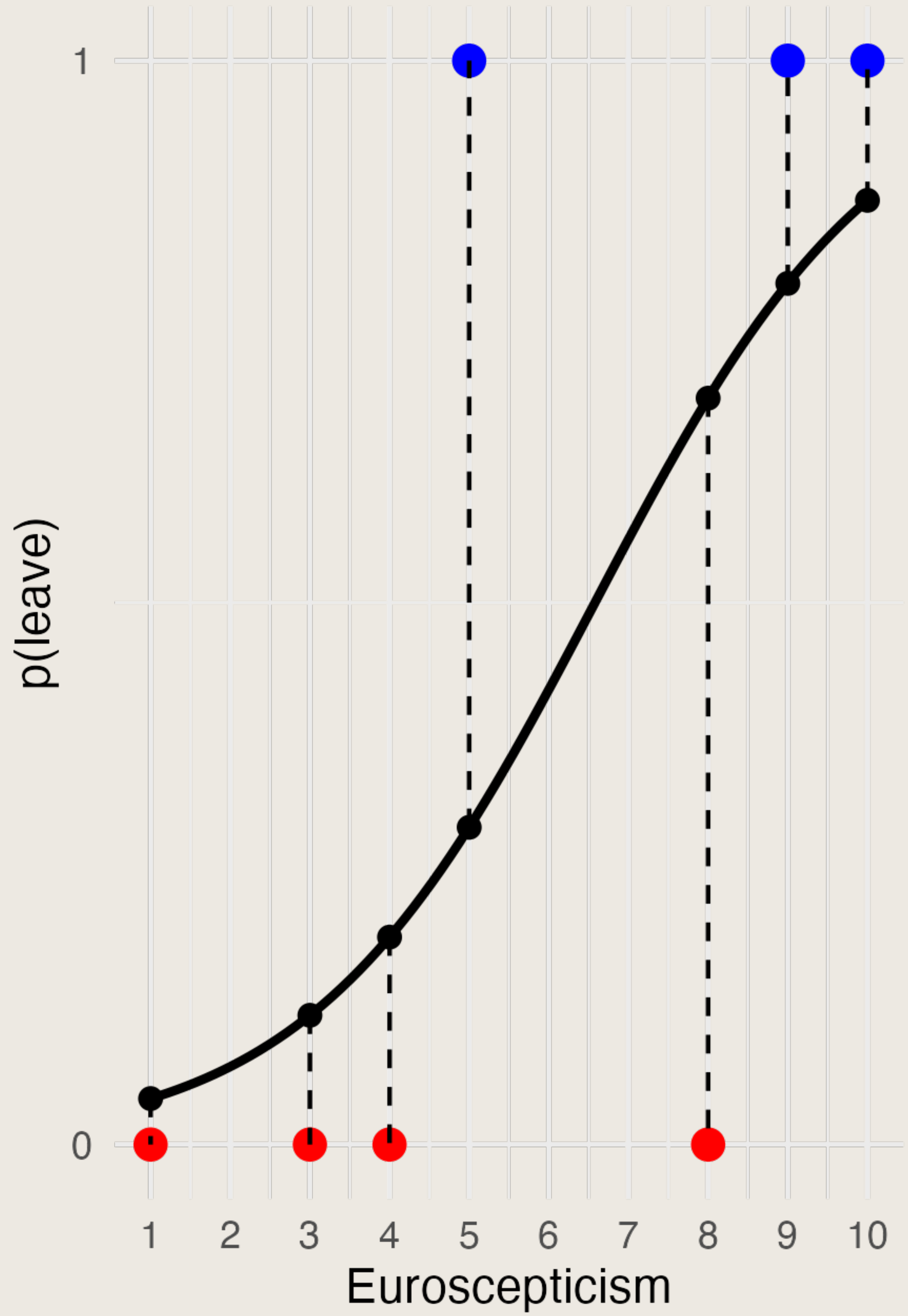
Maximum Likelihood: Intuition

- Likelihood: Pr(Data | Model).



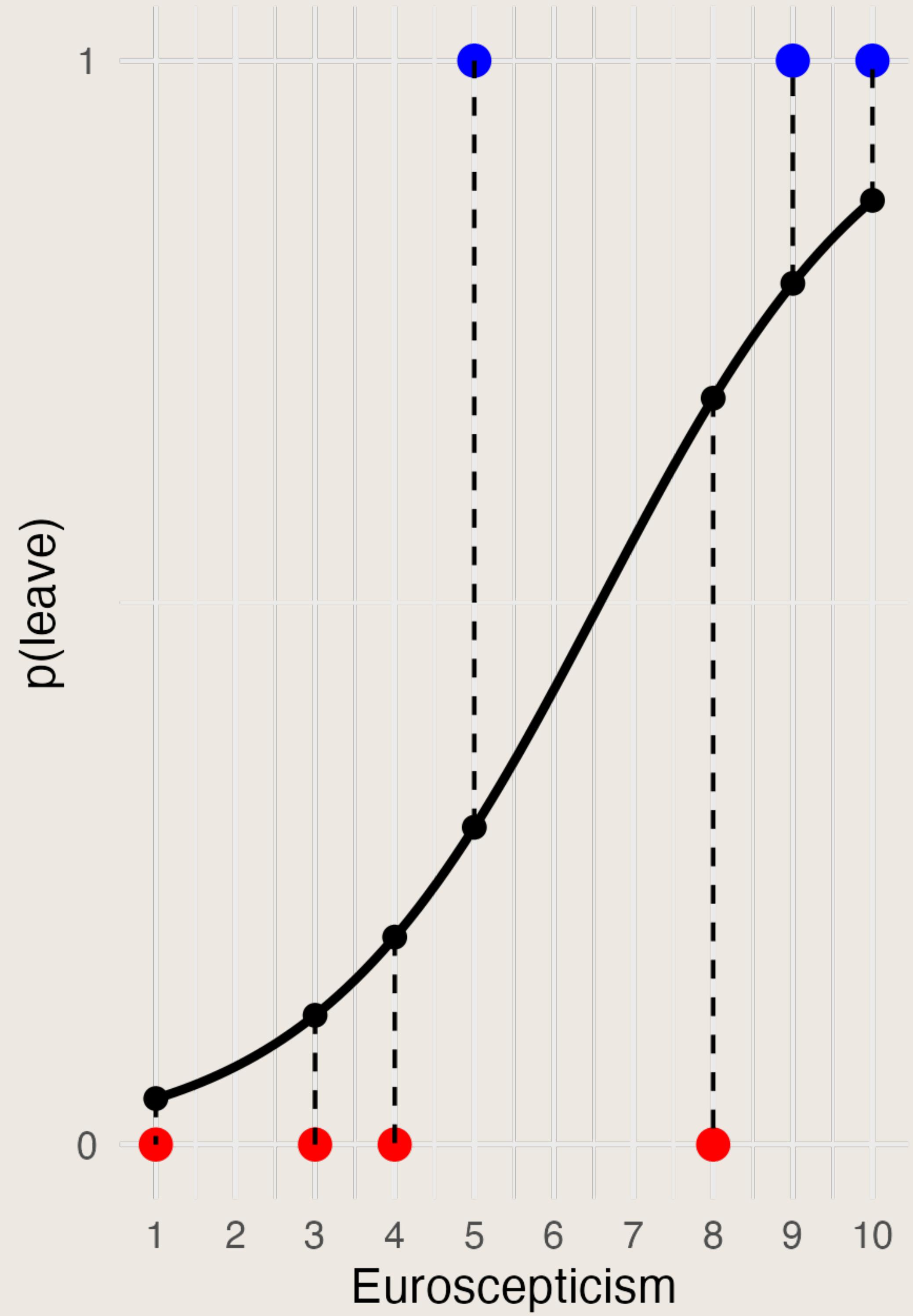
Maximum Likelihood: Intuition

- Likelihood: Pr(Data | Model).
- Given a logistic function with certain values of the coefficients α and β, what' the likelihood of observing the outcome for observation i?



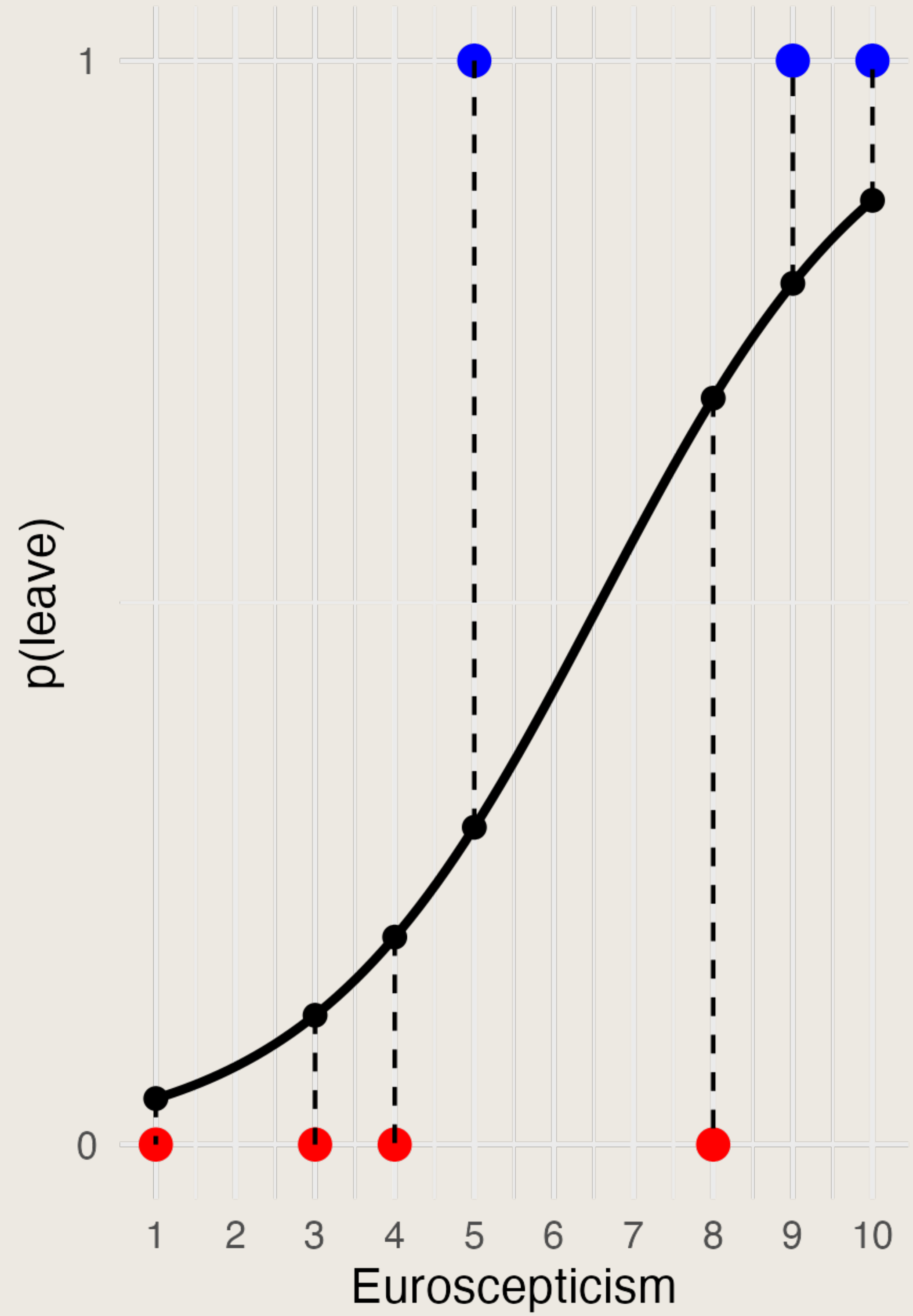
Maximum Likelihood: Intuition

- Likelihood: Pr(Data | Model).
- Given a logistic function with certain values of the coefficients α and β, what' the likelihood of observing the outcome for observation i?
- If Leave$_i$ = 1, it's Pr(Leave) = 1



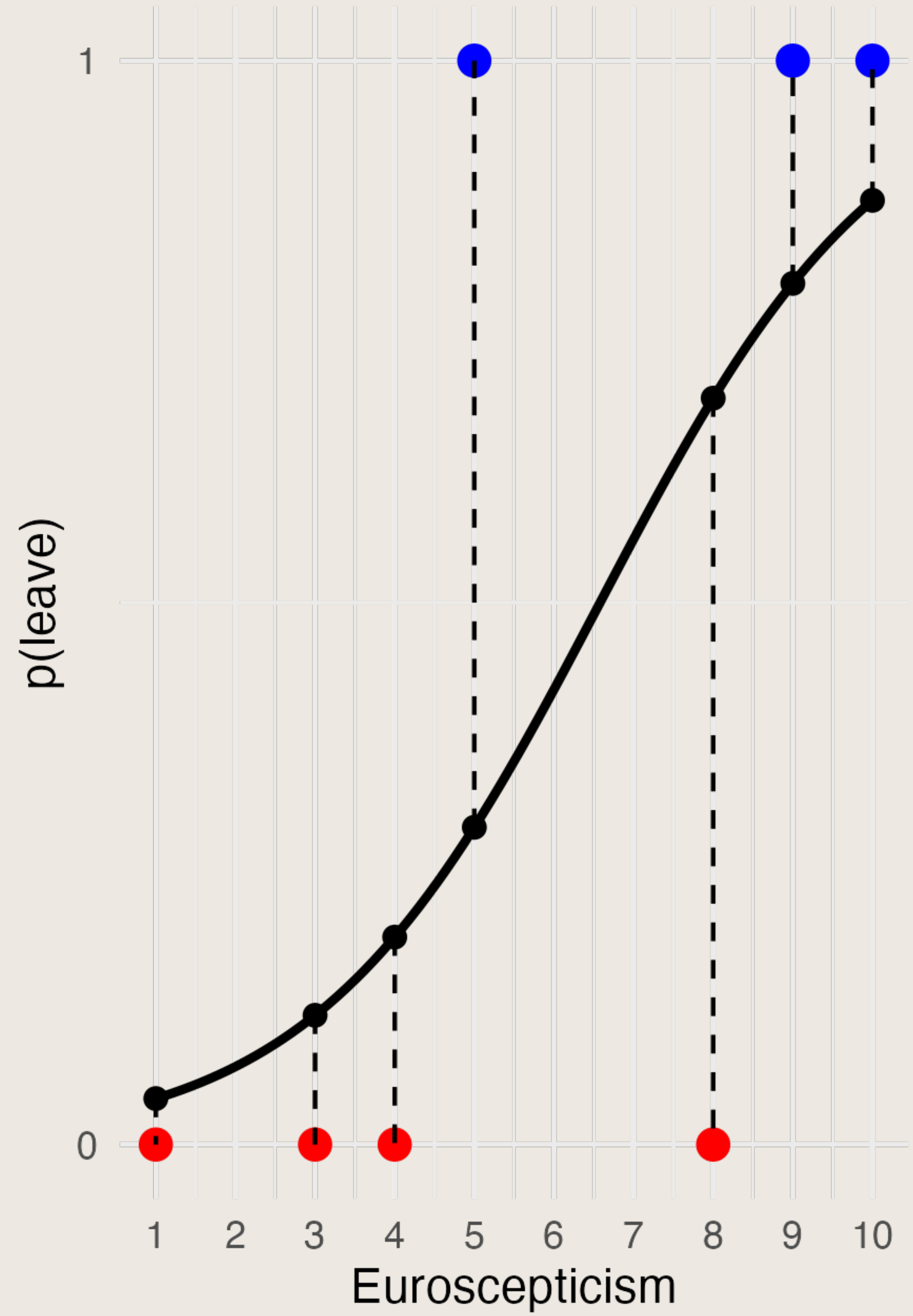
Maximum Likelihood: Intuition

- Likelihood: Pr(Data | Model).
- Given a logistic function with certain values of the coefficients α and β, what' the likelihood of observing the outcome for observation i?
- If $\text{Leave}_i = 1$, it's $\Pr(\text{Leave}) = 1$
- If $\text{Leave}_i = 0$, it's $1 - \Pr(\text{Leave}) = 1$



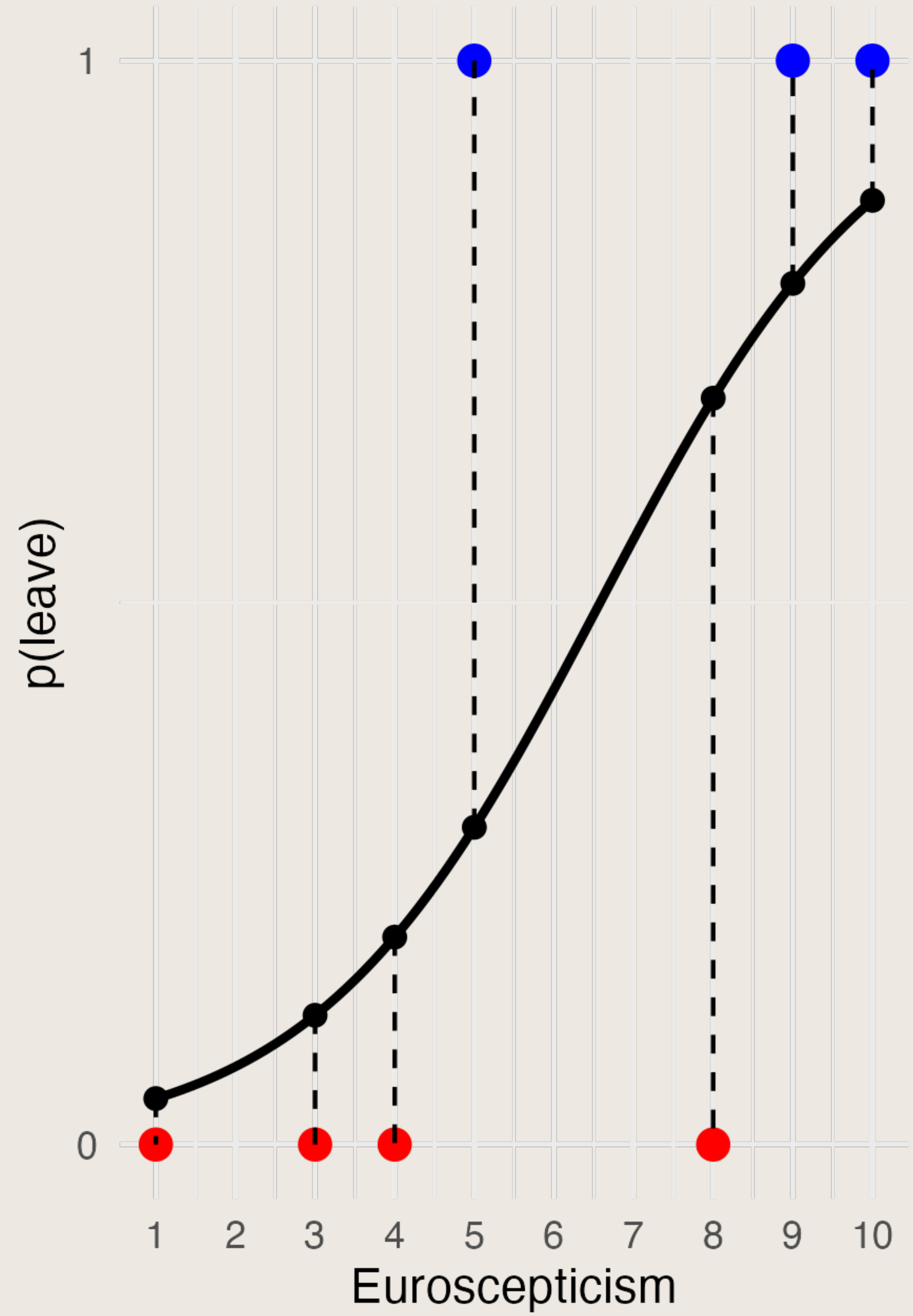
Maximum Likelihood: Intuition

- Likelihood: Pr(Data | Model).
- Given a logistic function with certain values of the coefficients α and β, what' the likelihood of observing the outcome for observation i?
- If Leave$_i$ = 1, it's Pr(Leave) = 1
- If Leave$_i$ = 0, it's 1 − Pr(Leave) = 1
- The overall likelihood of the data, given a model, is the probability of **jointly** observing all these outcomes → the product of the likelihoods.



Maximum Likelihood: Intuition

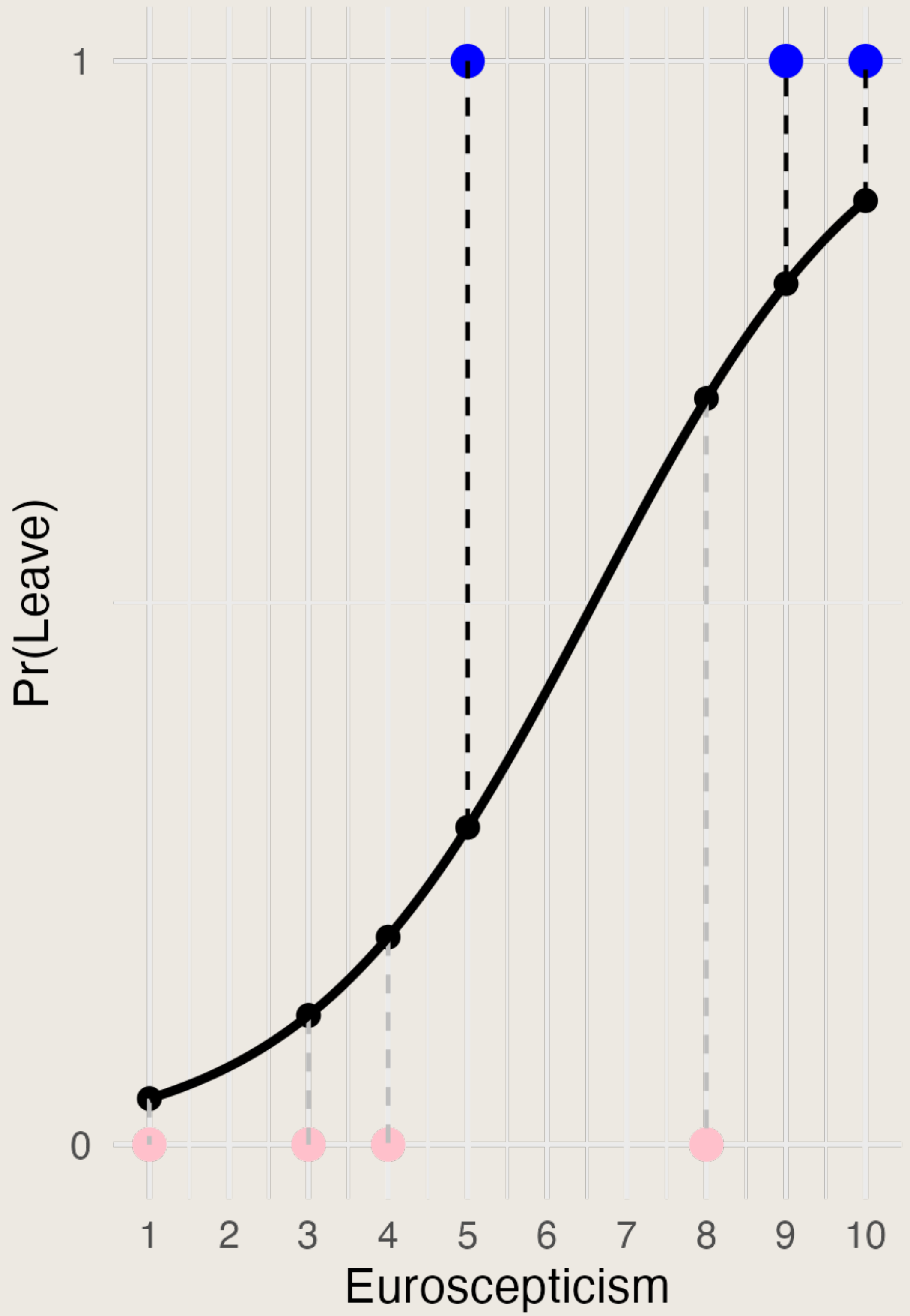
- Likelihood: Pr(Data | Model).
- Given a logistic function with certain values of the coefficients α and β, what' the likelihood of observing the outcome for observation i?
- If $\text{Leave}_i = 1$, it's $\text{Pr(Leave)} = 1$
- If $\text{Leave}_i = 0$, it's $1 - \text{Pr(Leave)} = 1$
- The overall likelihood of the data, given a model, is the probability of **jointly** observing all these outcomes → the product of the likelihoods.
- $\ell(\alpha, \beta) = \Pi_{Y_i=1}(\hat{p}_i)\ \Pi_{Y_i=0}(1 - \hat{p}_i)$



Maximum Likelihood: Intuition

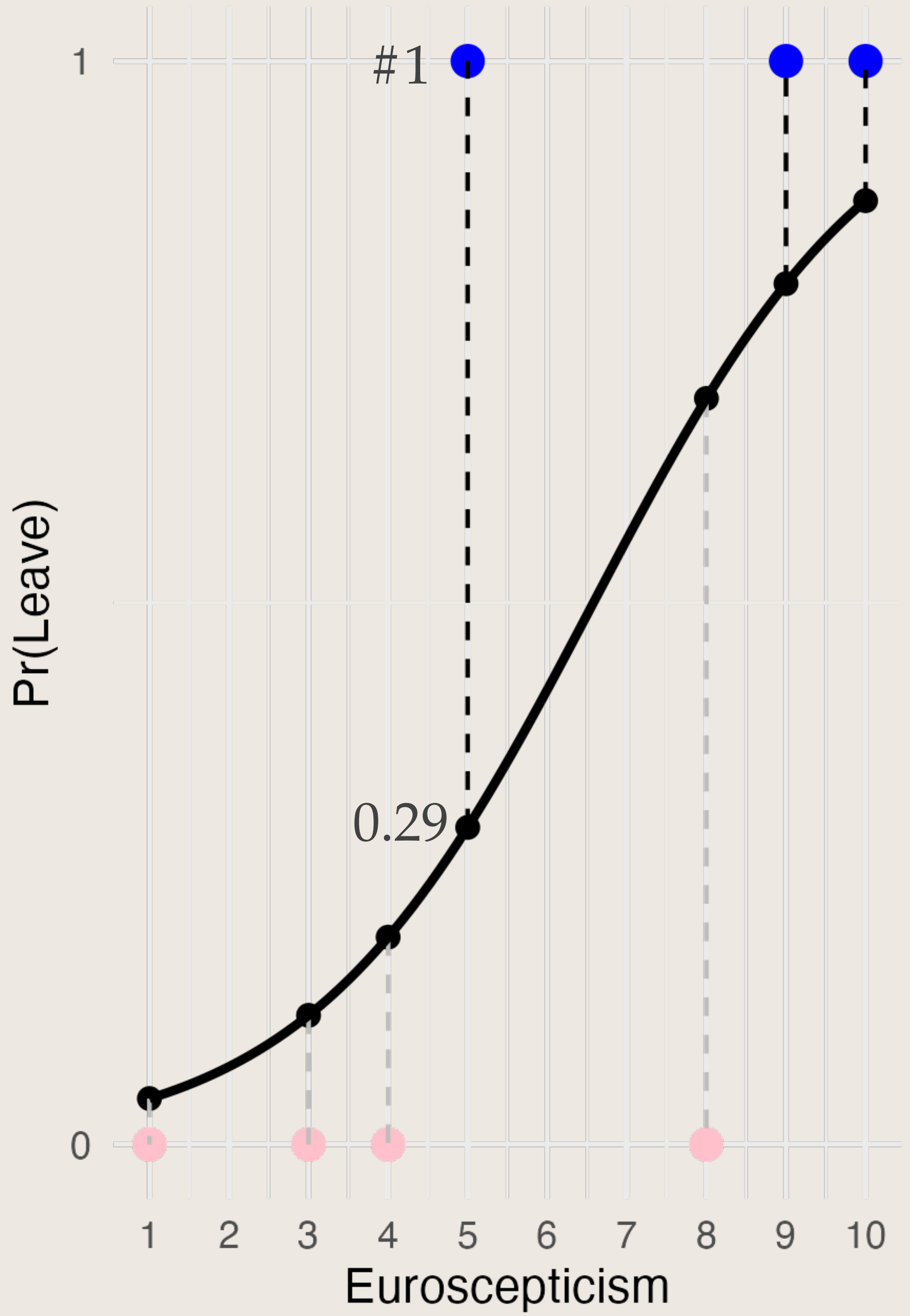
Maximum Likelihood: Intuition

R#	X	Y	Likelihood



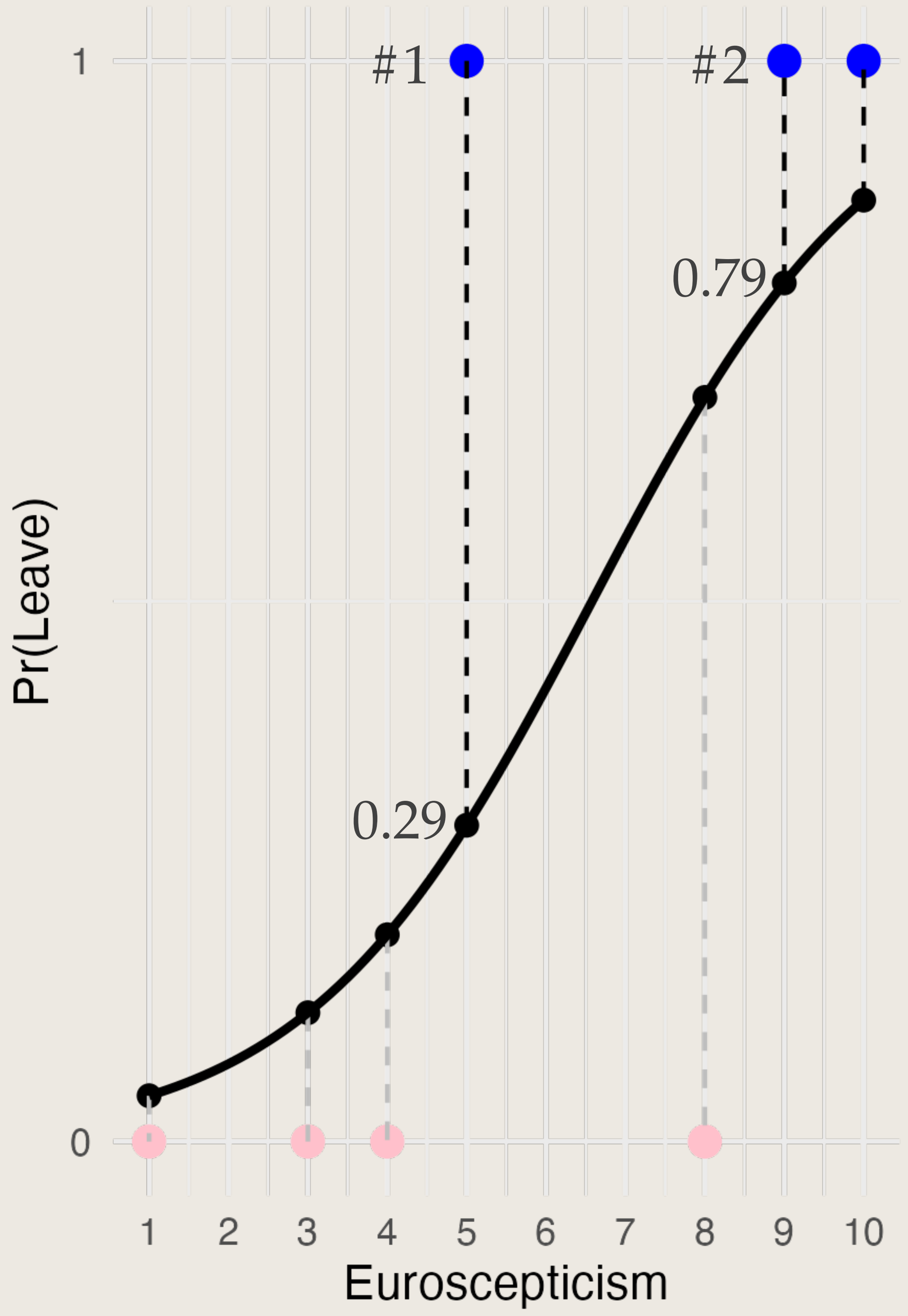
Maximum Likelihood: Intuition

R#	**X**	**Y**	**Likelihood**
1	5	1	0.29



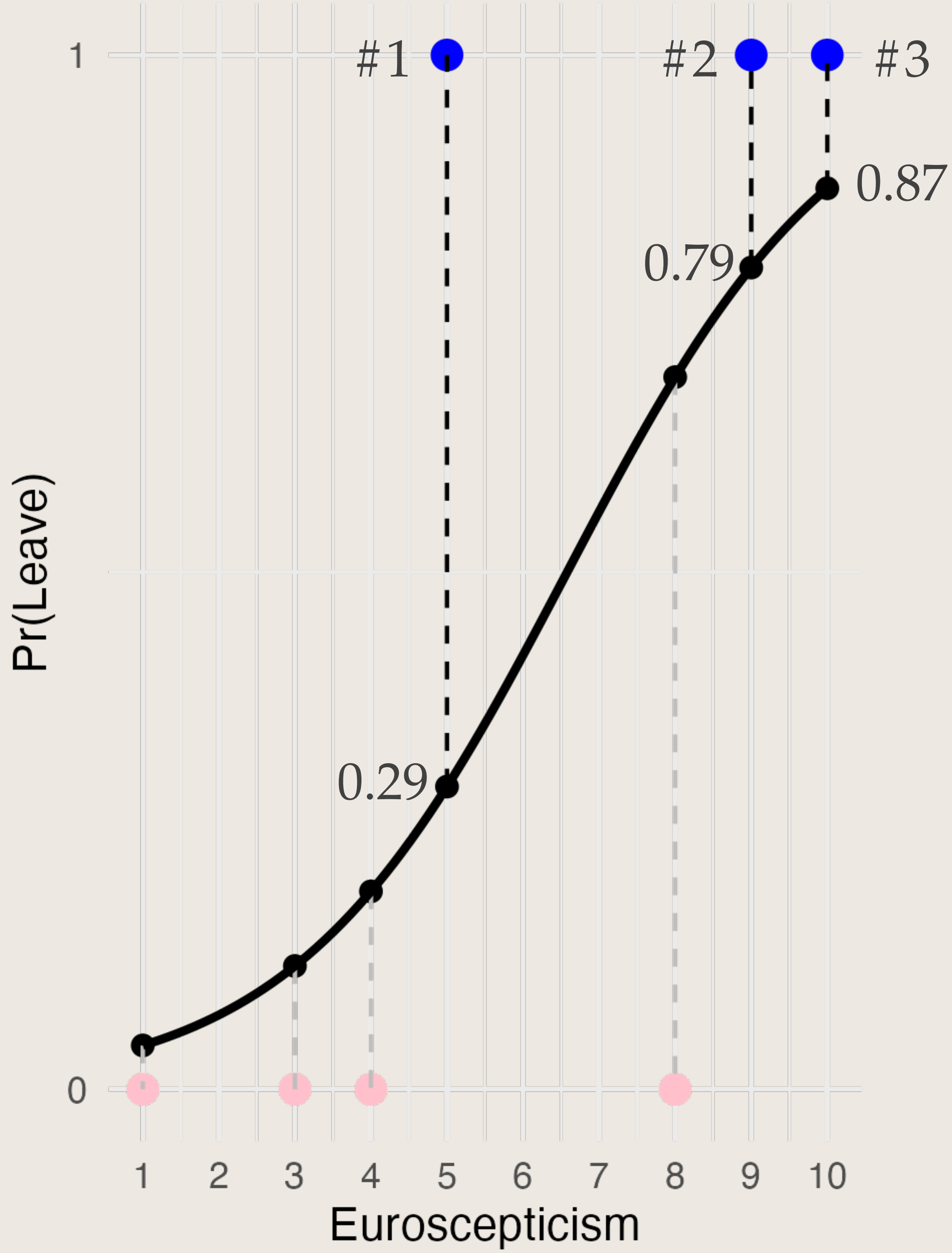
Maximum Likelihood: Intuition

R#	**X**	**Y**	**Likelihood**
1	5	1	0.29
2	9	1	0.79



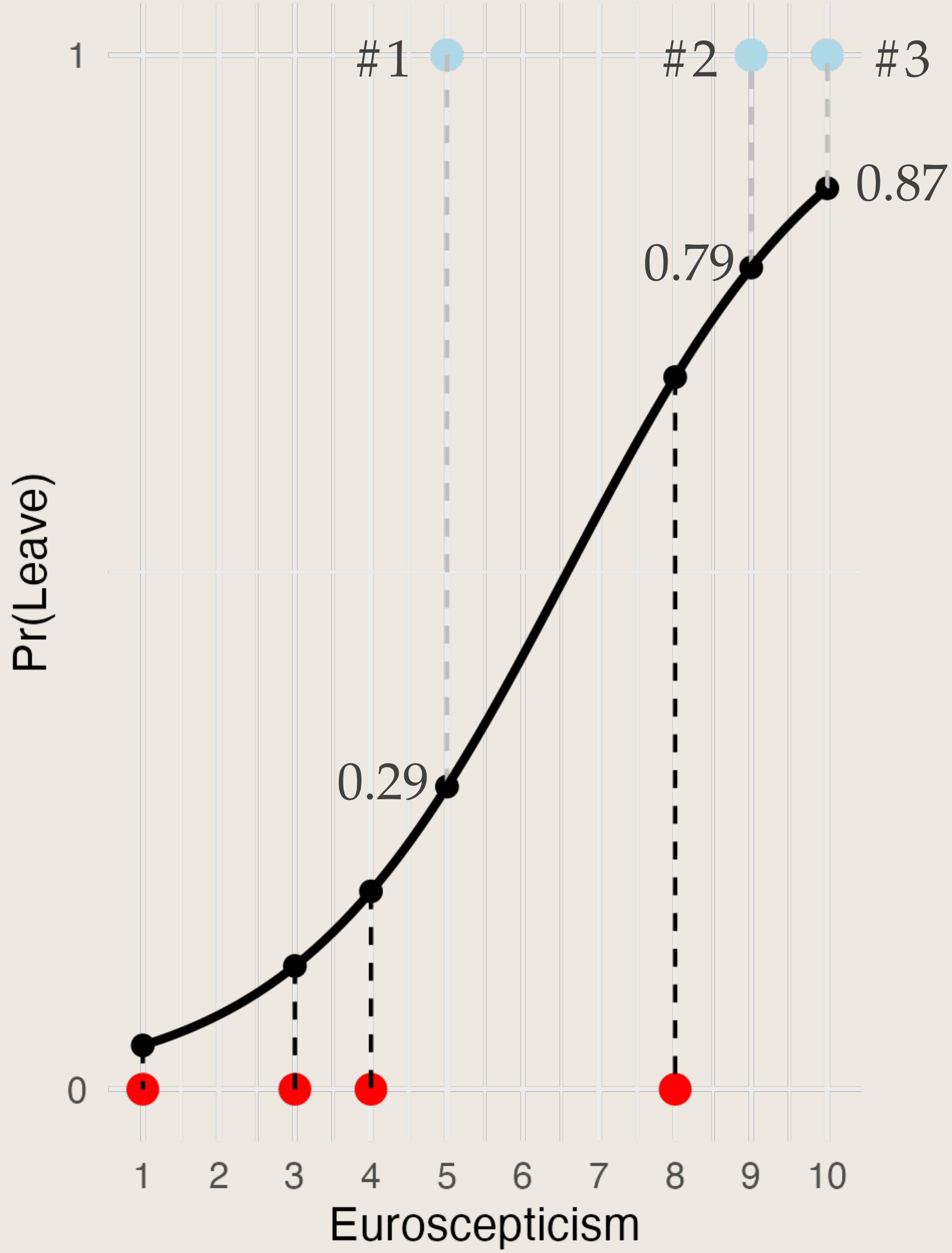
Maximum Likelihood: Intuition

R#	X	Y	Likelihood
1	5	1	0.29
2	9	1	0.79
3	10	1	0.87



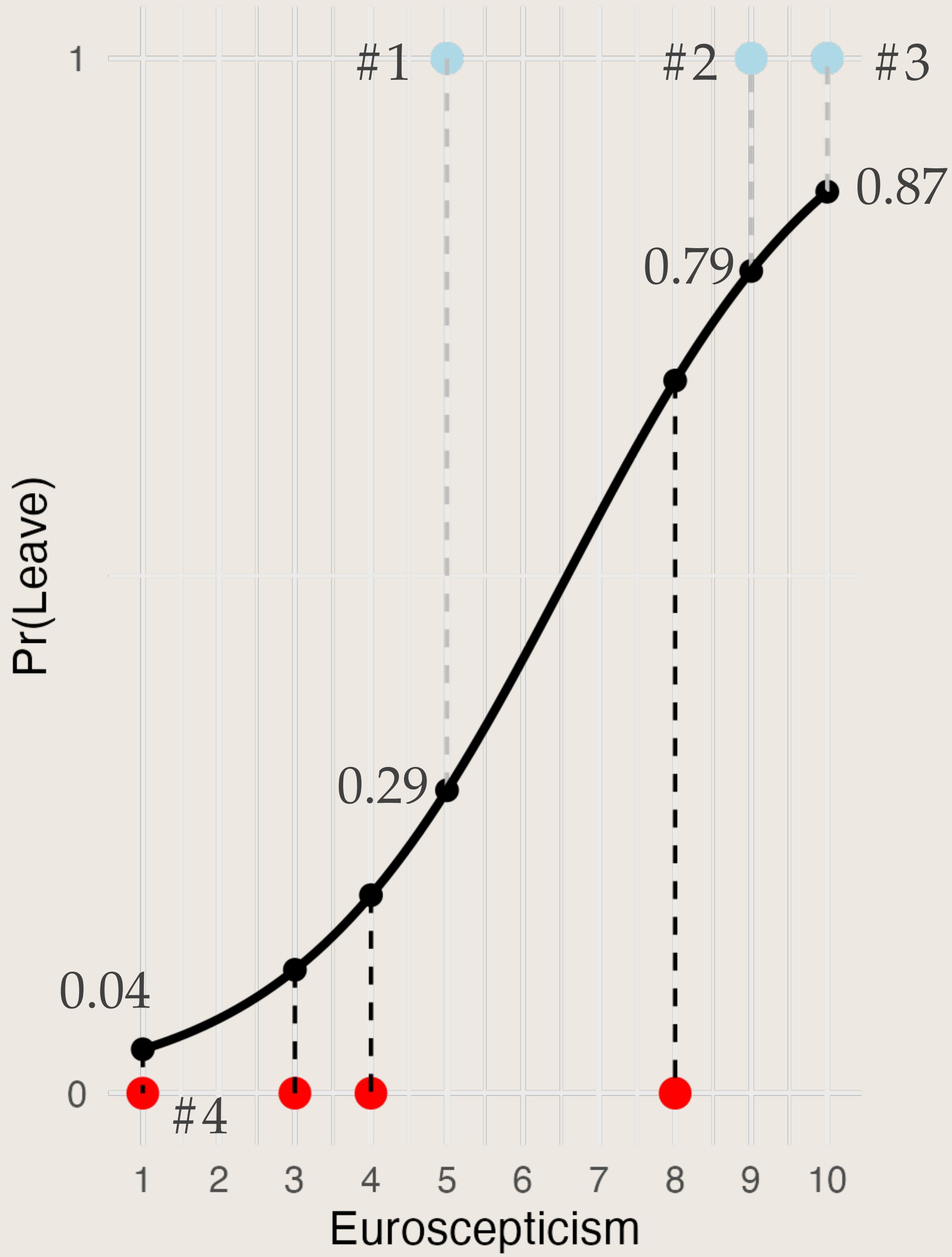
Maximum Likelihood: Intuition

R#	**X**	**Y**	**Likelihood**
1	5	1	0.29
2	9	1	0.79
3	10	1	0.87



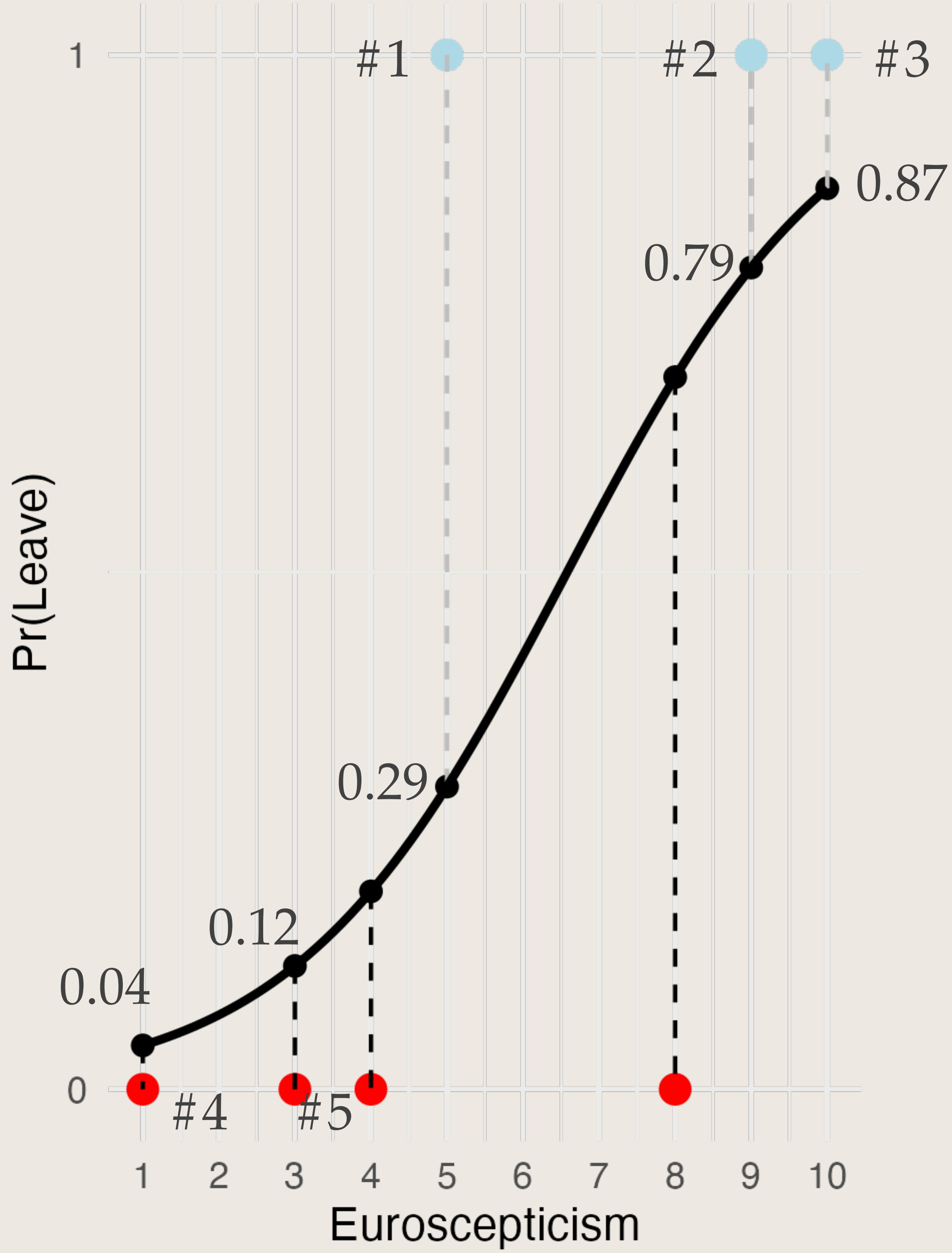
Maximum Likelihood: Intuition

R#	**X**	**Y**	**Likelihood**
1	5	1	0.29
2	9	1	0.79
3	10	1	0.87
4	1	0	1–0.04 = 0.96



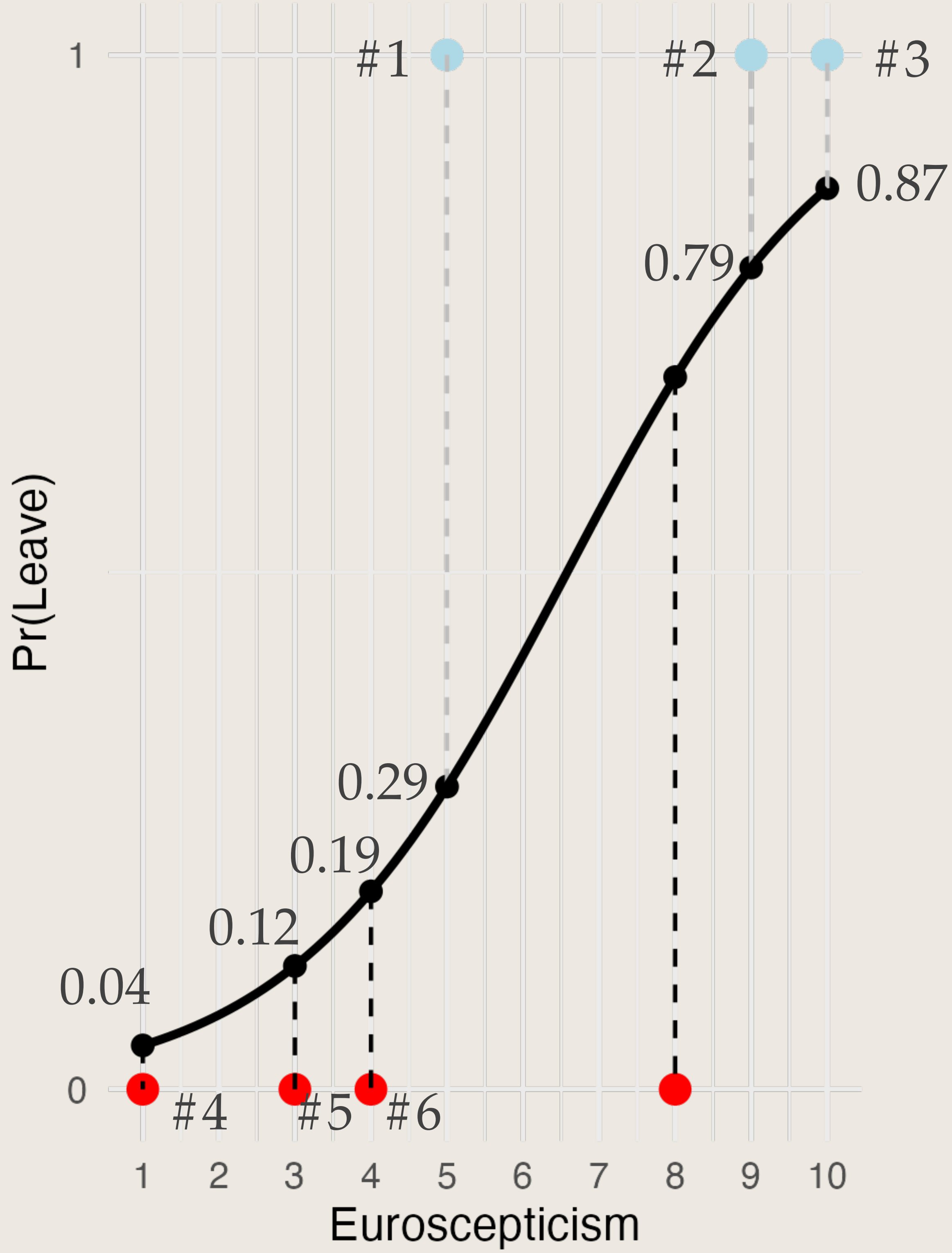
Maximum Likelihood: Intuition

R#	**X**	**Y**	**Likelihood**
1	5	1	0.29
2	9	1	0.79
3	10	1	0.87
4	1	0	1–0.04 = 0.96
5	3	0	1–0.12 = 0.88



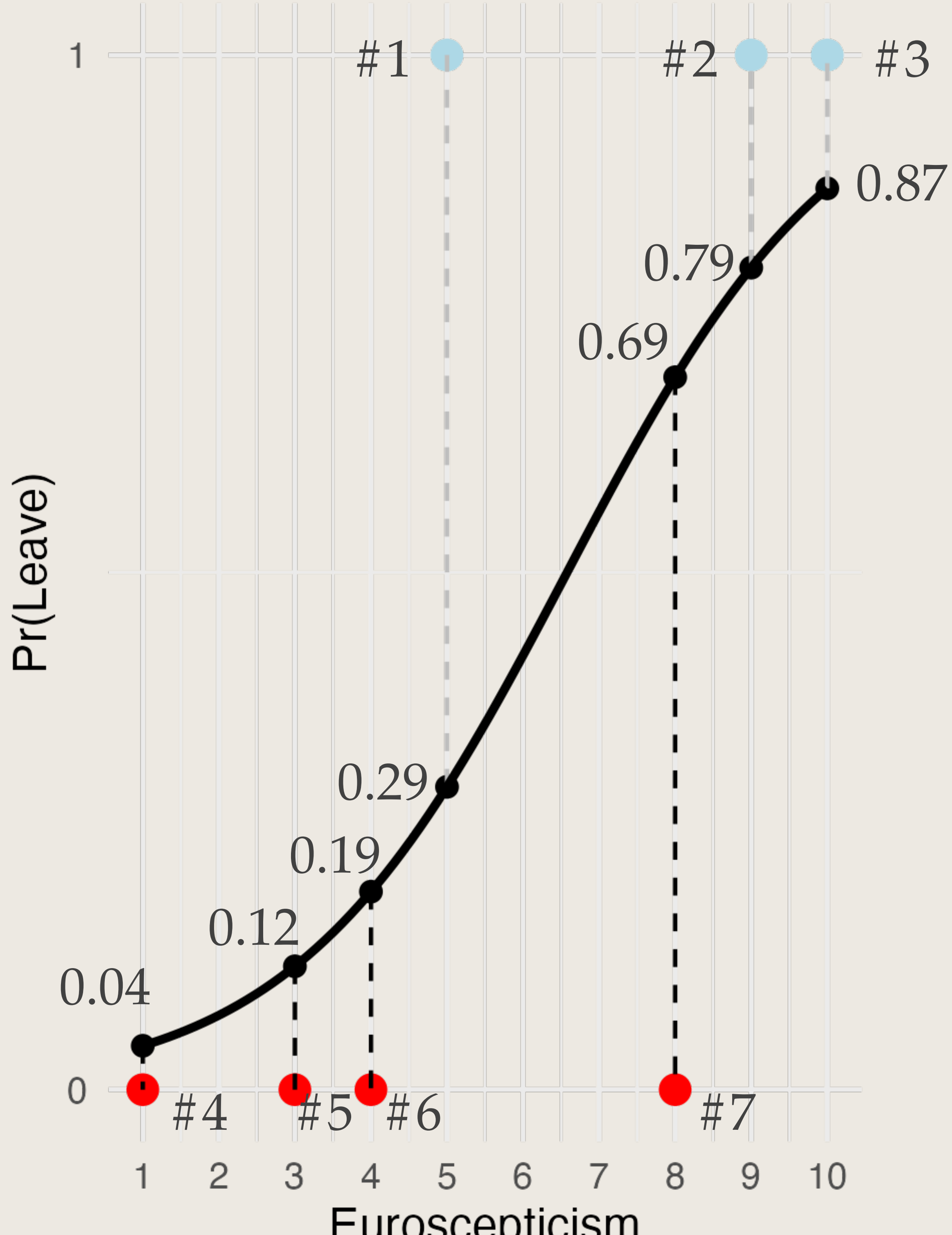
Maximum Likelihood: Intuition

R#	**X**	**Y**	**Likelihood**
1	5	1	0.29
2	9	1	0.79
3	10	1	0.87
4	1	0	1–0.04 = 0.96
5	3	0	1–0.12 = 0.88
6	4	0	1–0.19 = 0.81



Maximum Likelihood: Intuition

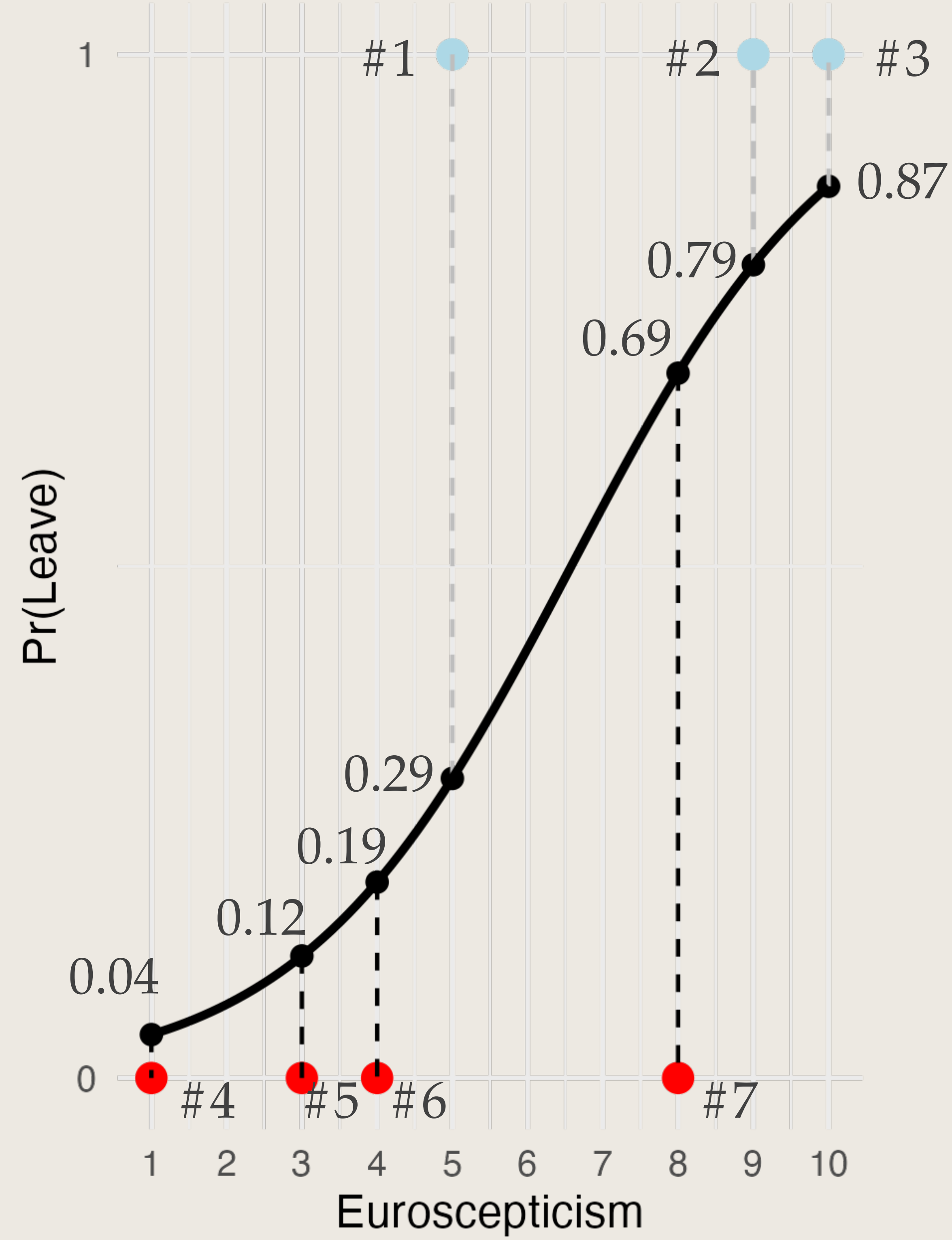
R#	X	Y	Likelihood
1	5	1	0.29
2	9	1	0.79
3	10	1	0.87
4	1	0	1–0.04 = 0.96
5	3	0	1–0.12 = 0.88
6	4	0	1–0.19 = 0.81
7	8	0	1–0.69 = 0.31



Maximum Likelihood: Intuition

R#	**X**	**Y**	**Likelihood**
1	5	1	0.29
2	9	1	0.79
3	10	1	0.87
4	1	0	1–0.04 = 0.96
5	3	0	1–0.12 = 0.88
6	4	0	1–0.19 = 0.81
7	8	0	1–0.69 = 0.31
Likelihood of the model			0.043

$$0.29 \times 0.79 \times 0.87 \times 0.96 \times 0.88 \times 0.81 \times 0.31$$



Maximum Likelihood: Intuition

Maximum Likelihood: Intuition

* No closed-form solution: R will go over many possible sigmoids, and find the coefficients that return the highest joint likelihood ℓ of the data — hence, **maximum likelihood**.

Maximum Likelihood: Intuition

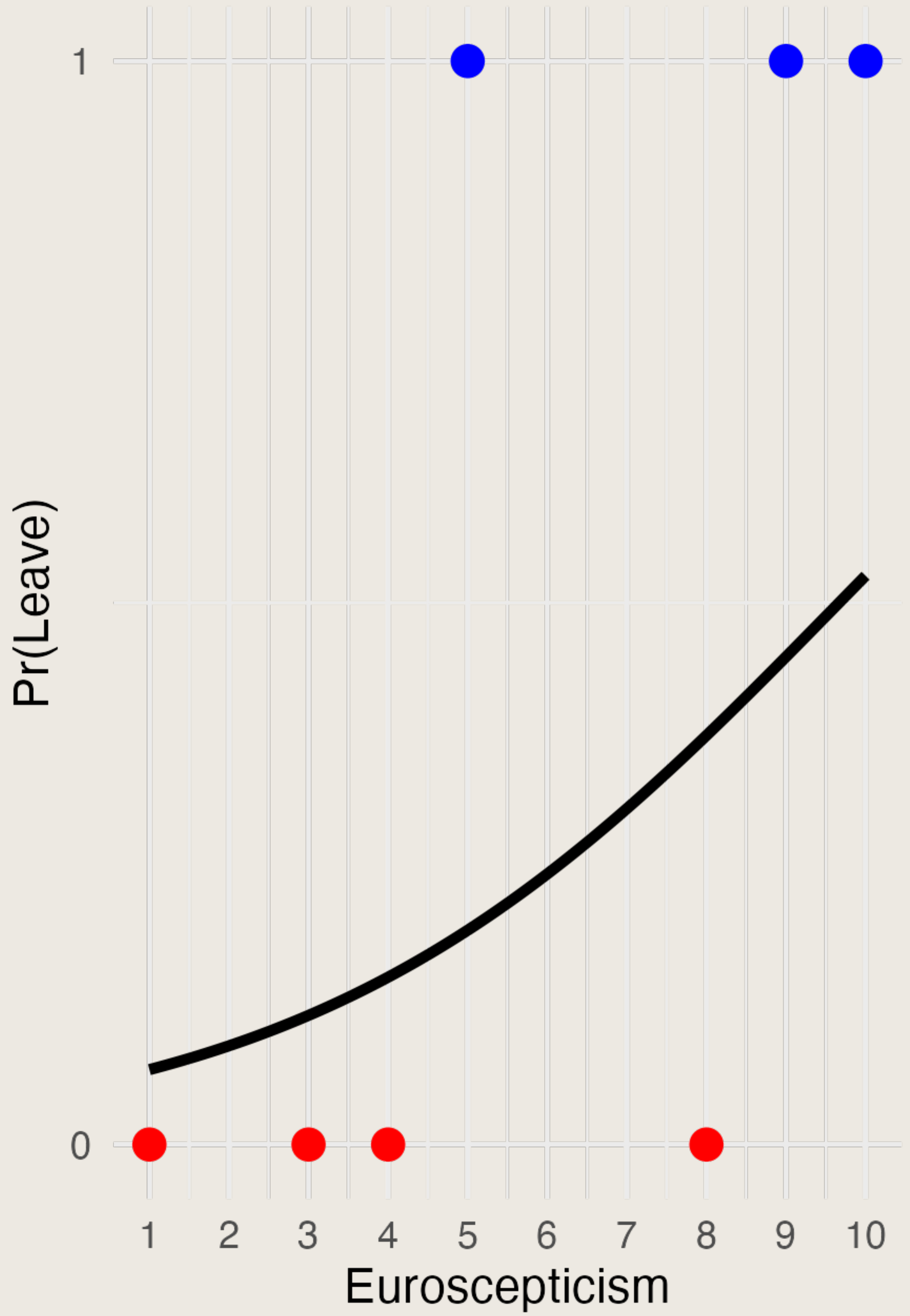
- No closed-form solution: R will go over many possible sigmoids, and find the coefficients that return the highest joint likelihood ℓ of the data — hence, **maximum likelihood**.
- In fact, for ease of computation, it will seek the one with the **maximum log-likelihood** (LL).

Maximum Likelihood: Intuition

- No closed-form solution: R will go over many possible sigmoids, and find the coefficients that return the highest joint likelihood ℓ of the data — hence, **maximum likelihood**.
- In fact, for ease of computation, it will seek the one with the **maximum log-likelihood** (LL).
- Nothing to worry about: the maximum LL occurs for the same coefficients as the maximum of ℓ.

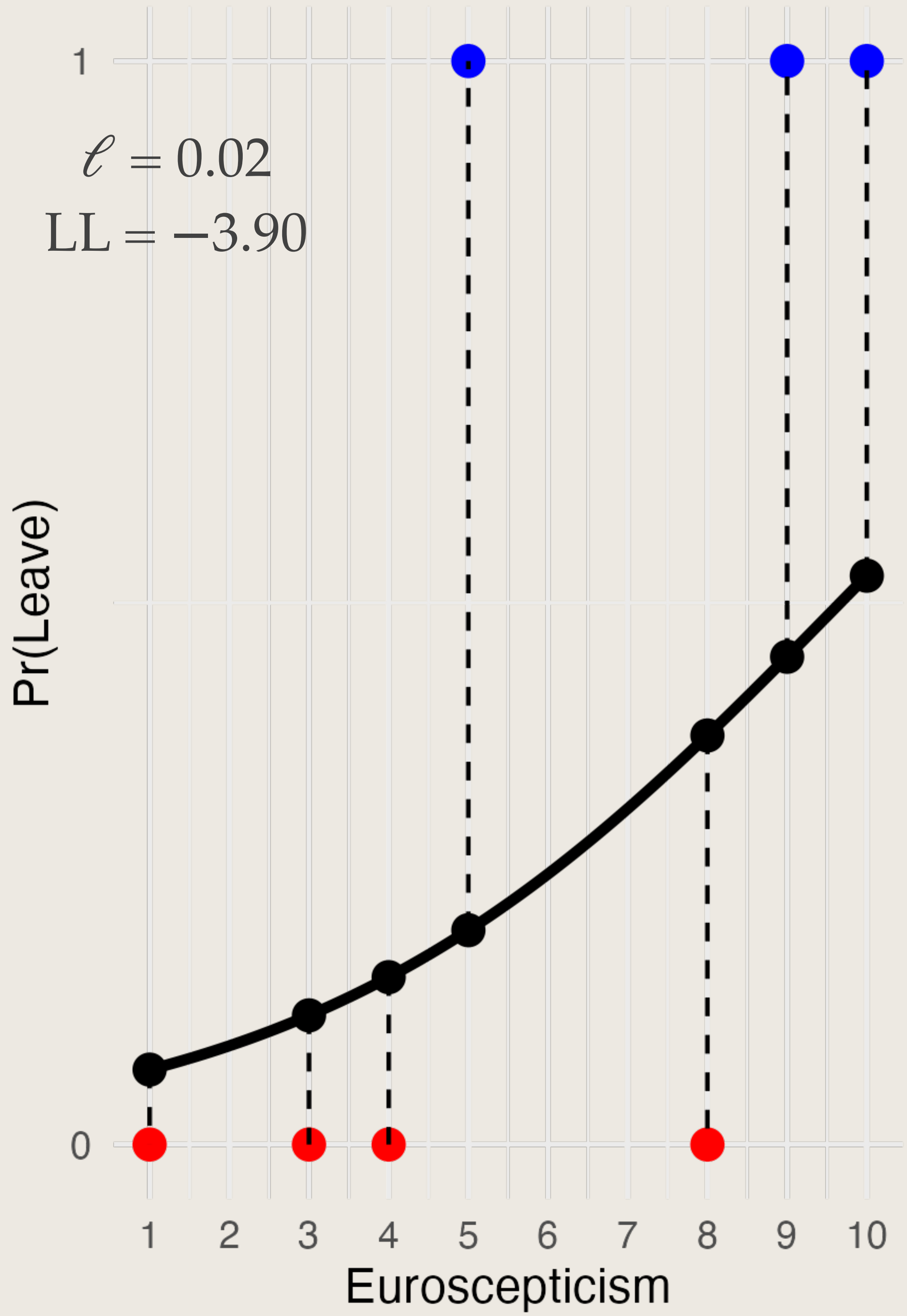
Maximum Likelihood: Intuition

- No closed-form solution: R will go over many possible sigmoids, and find the coefficients that return the highest joint likelihood ℓ of the data — hence, **maximum likelihood**.
- In fact, for ease of computation, it will seek the one with the **maximum log-likelihood** (LL).
- Nothing to worry about: the maximum LL occurs for the same coefficients as the maximum of ℓ.



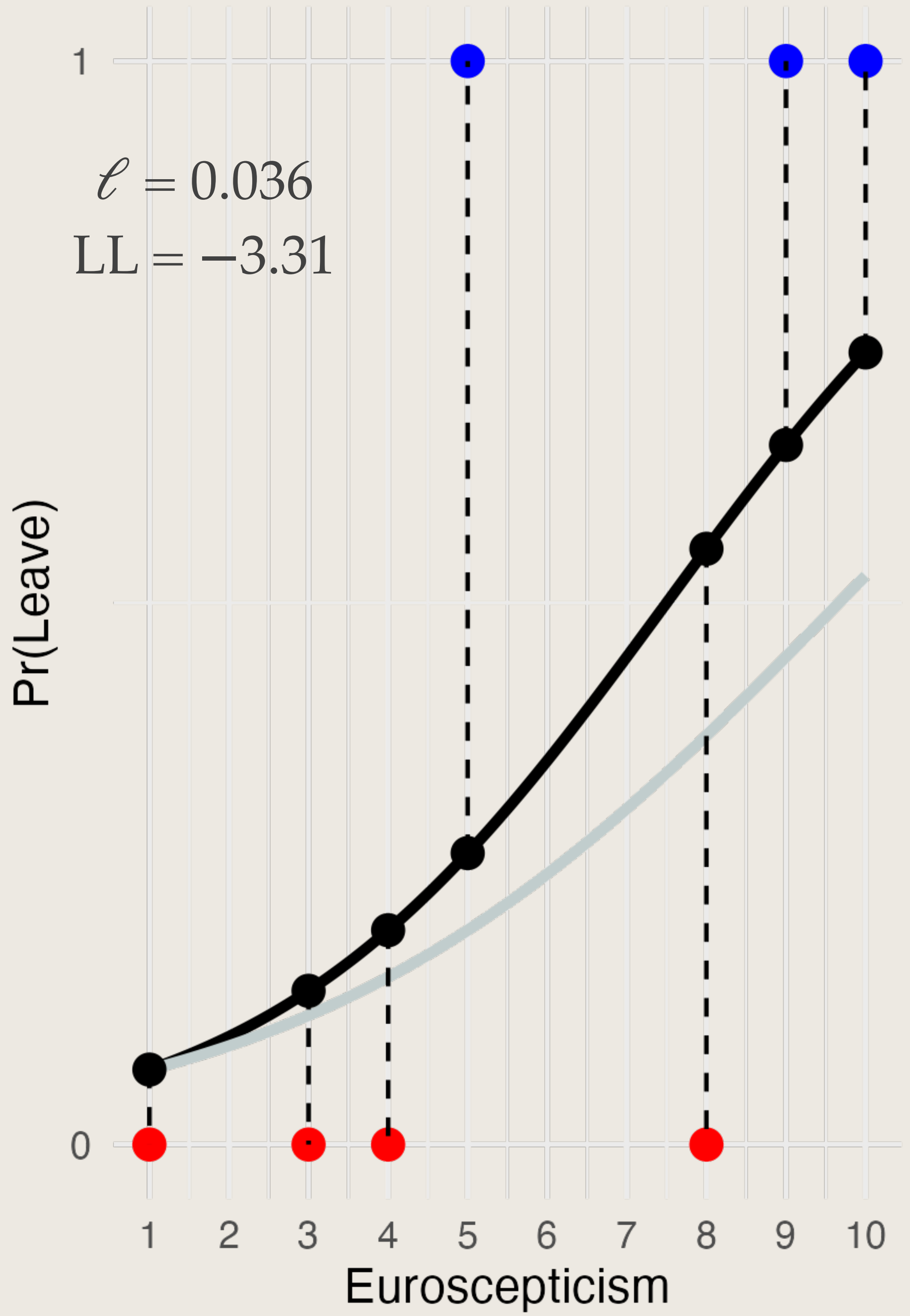
Maximum Likelihood: Intuition

- No closed-form solution: R will go over many possible sigmoids, and find the coefficients that return the highest joint likelihood ℓ of the data — hence, **maximum likelihood**.
- In fact, for ease of computation, it will seek the one with the **maximum log-likelihood** (LL).
- Nothing to worry about: the maximum LL occurs for the same coefficients as the maximum of ℓ.



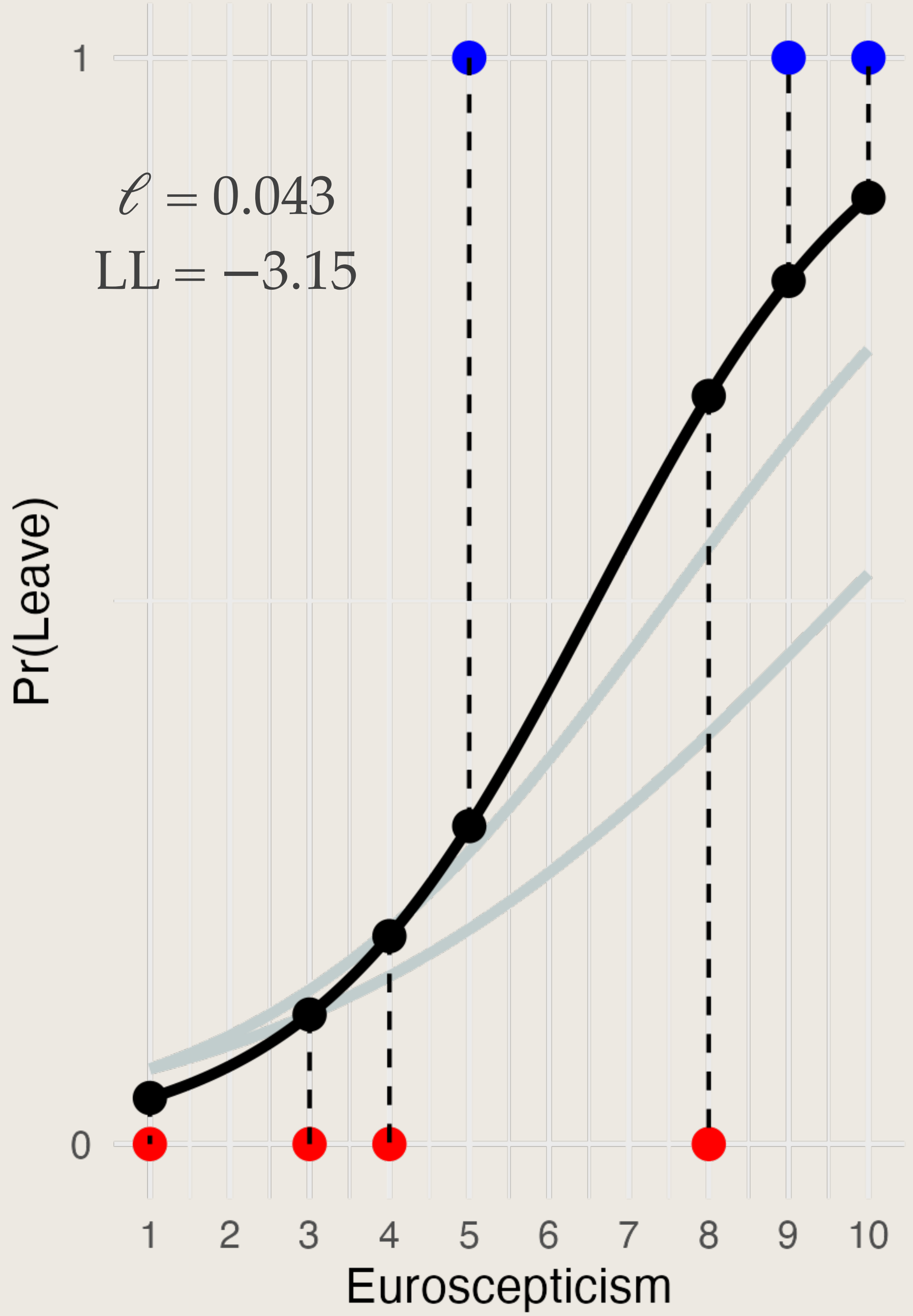
Maximum Likelihood: Intuition

- No closed-form solution: R will go over many possible sigmoids, and find the coefficients that return the highest joint likelihood ℓ of the data — hence, **maximum likelihood**.
- In fact, for ease of computation, it will seek the one with the **maximum log-likelihood** (LL).
- Nothing to worry about: the maximum LL occurs for the same coefficients as the maximum of ℓ.



Maximum Likelihood: Intuition

* No closed-form solution: R will go over many possible sigmoids, and find the coefficients that return the highest joint likelihood ℓ of the data — hence, **maximum likelihood**.
* In fact, for ease of computation, it will seek the one with the **maximum log-likelihood** (LL).
* Nothing to worry about: the maximum LL occurs for the same coefficients as the maximum of ℓ.



Goodness of fit

Goodness of fit

- No R^2, for the same reason that we can't do OLS: squared residuals are all ∞.

Goodness of fit

- No R^2, for the same reason that we can't do OLS: squared residuals are all ∞.
- Software normally returns the Log-Likelihood (LL), and something called the Akaike Information Criterion. Neither is particularly useful.

```
> stargazer(model, type = "text", single.row = TRUE)

=============================================
                      Dependent variable:
                  ---------------------------
                             Leave
---------------------------------------------
Euroscepticism          1.020*** (0.143)
trustMPs                -0.353** (0.172)
genderFemale              0.164 (0.513)
Constant               -5.655*** (1.031)
---------------------------------------------
Observations                  199
Log Likelihood              -50.830
Akaike Inf. Crit.           109.660
=============================================
Note:              *p<0.1; **p<0.05; ***p<0.01
>
```

Goodness of fit

- No R^2, for the same reason that we can't do OLS: squared residuals are all ∞.
- Software normally returns the Log-Likelihood (LL), and something called the Akaike Information Criterion. Neither is particularly useful.
- Pseudo-R^2 is a bit better: compares the LL of the model with the LL of a model without independent variables.

```
> stargazer(model, type = "text", single.row = TRUE)

==============================================
                      Dependent variable:
                  ----------------------------
                             Leave
----------------------------------------------
Euroscepticism          1.020*** (0.143)
trustMPs                -0.353** (0.172)
genderFemale             0.164 (0.513)
Constant                -5.655*** (1.031)
----------------------------------------------
Observations                  199
Log Likelihood              -50.830
Akaike Inf. Crit.           109.660
==============================================
Note:              *p<0.1; **p<0.05; ***p<0.01
>
```

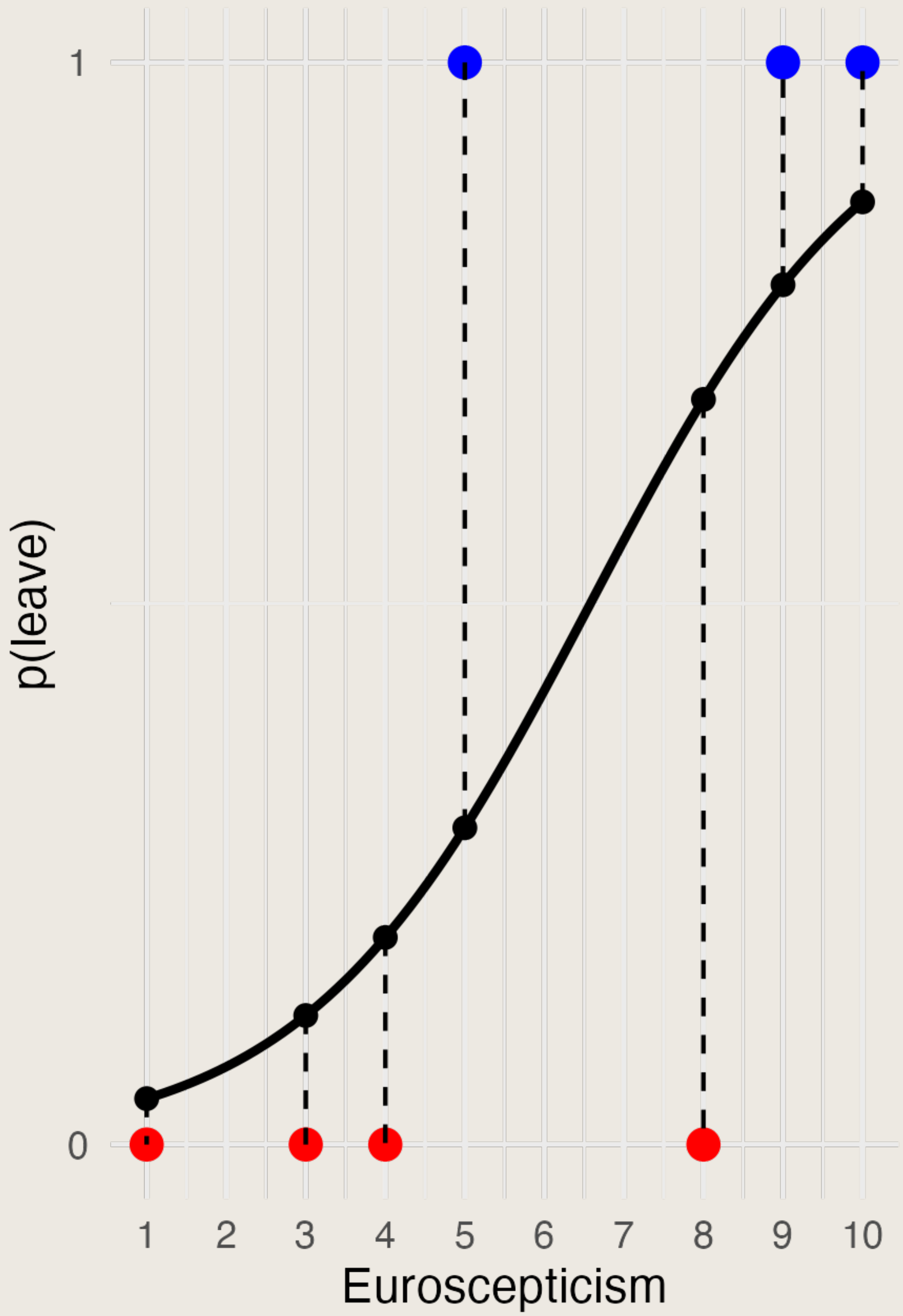
Goodness of fit

Goodness of fit

Our Model

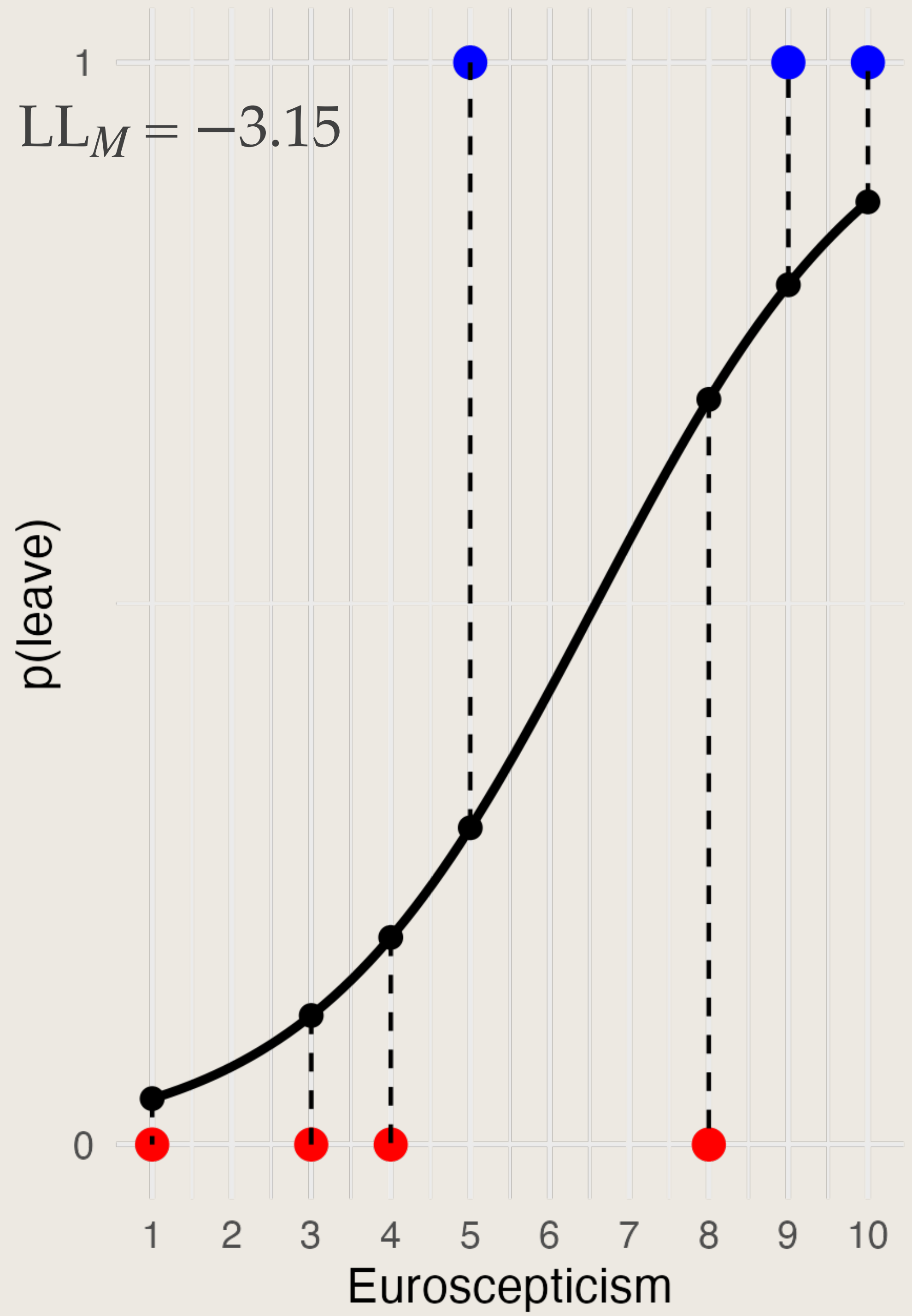
Goodness of fit

Our Model



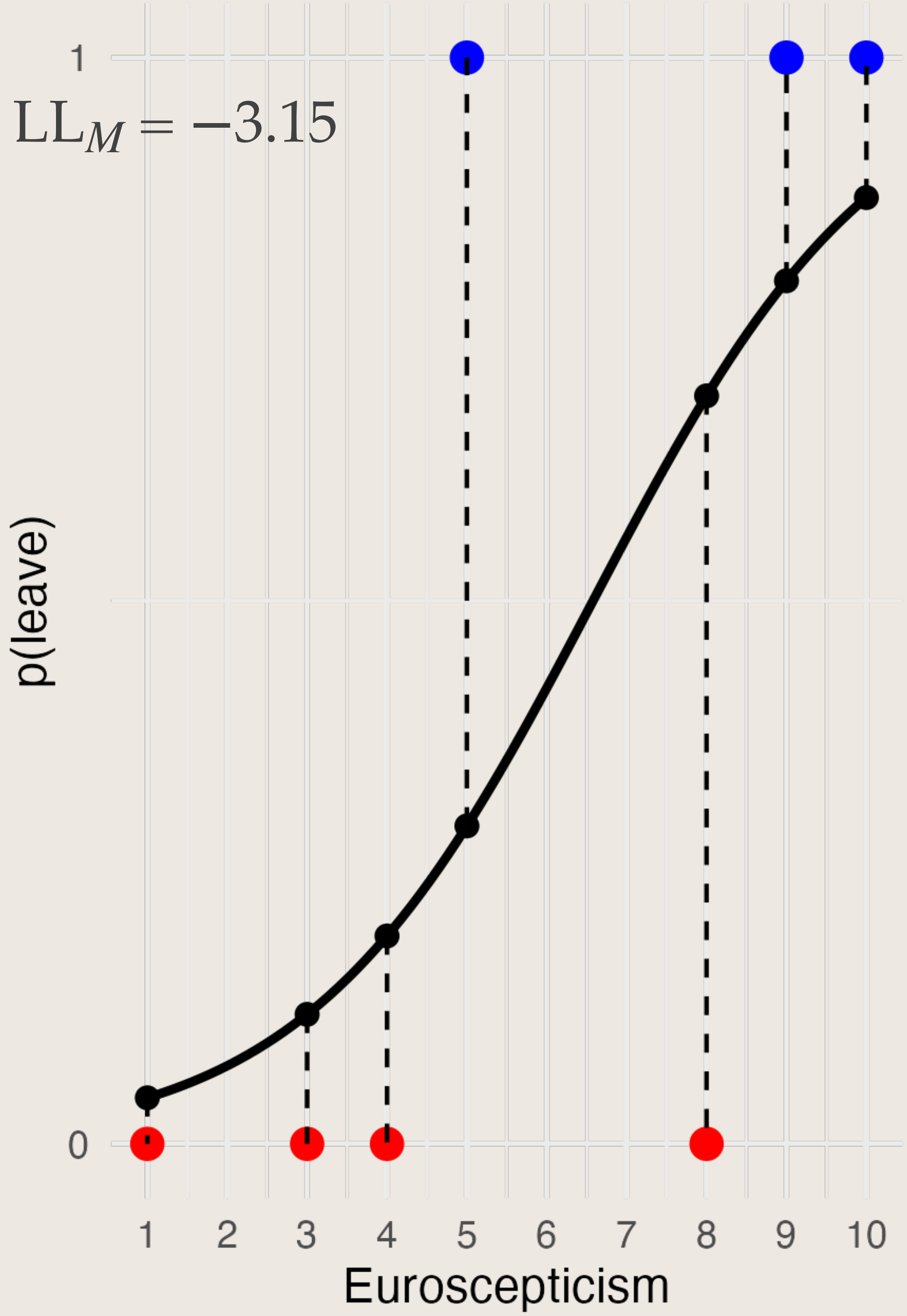
Goodness of fit

Our Model

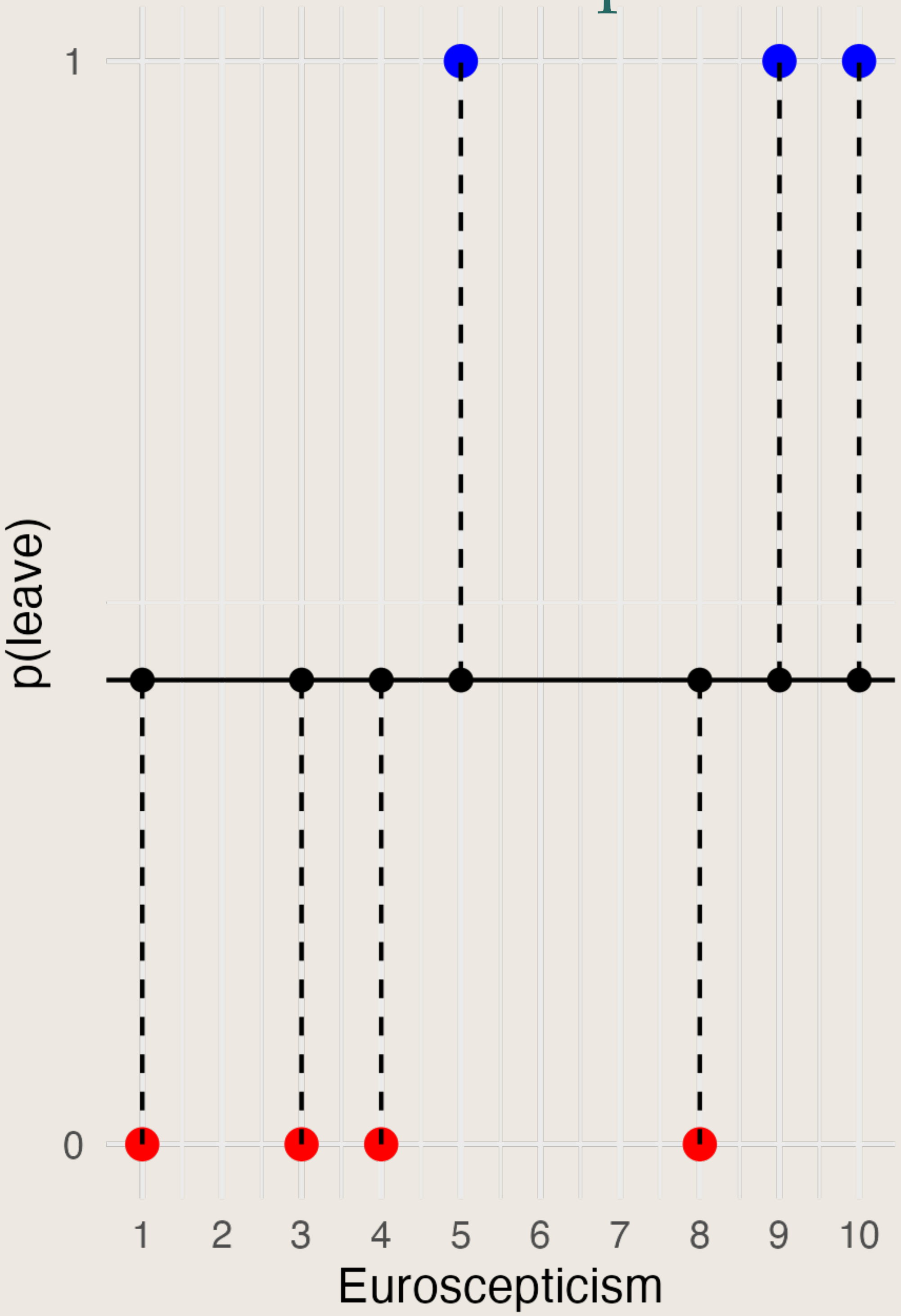


Goodness of fit

Our Model

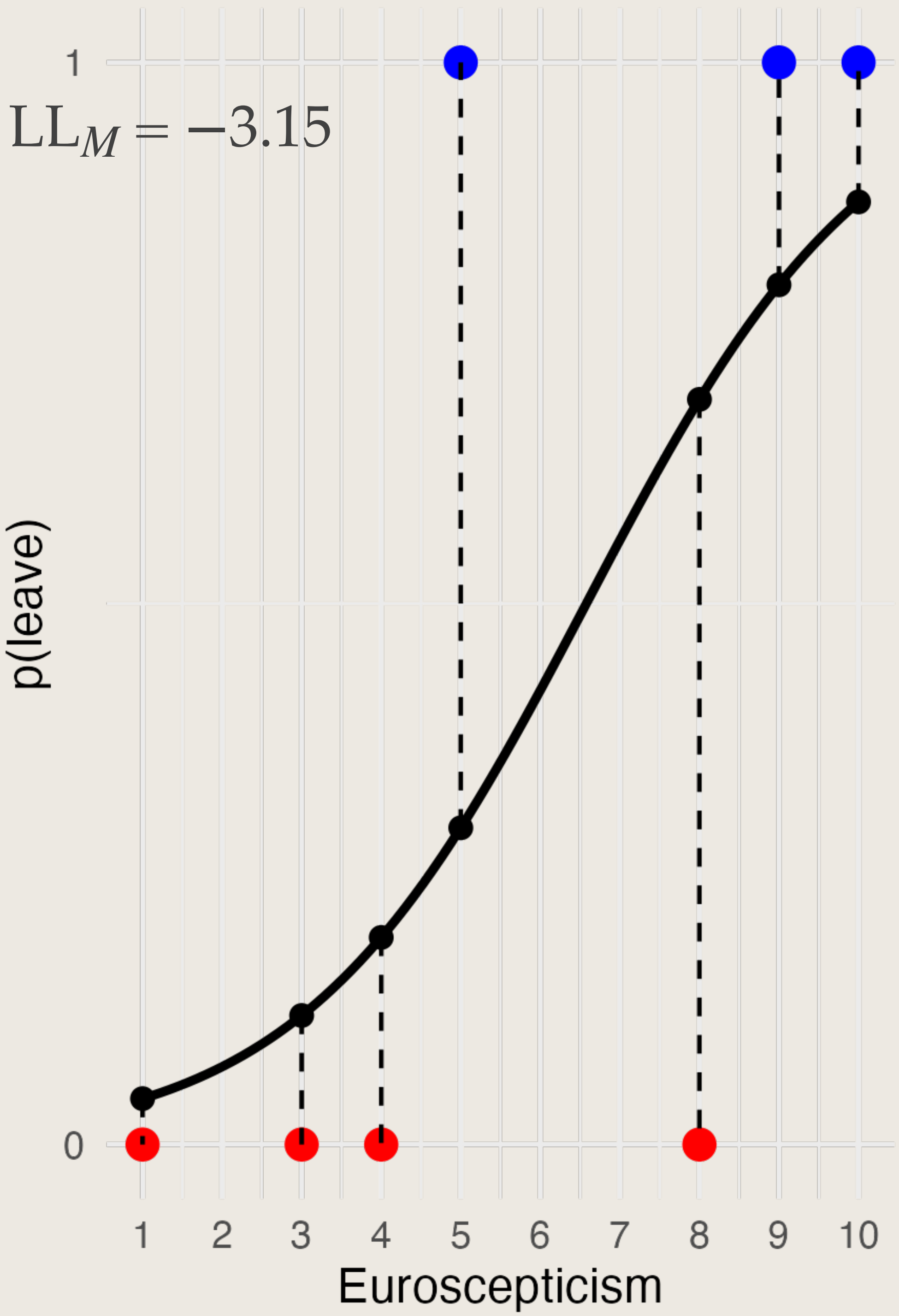


Best Guess without predictors

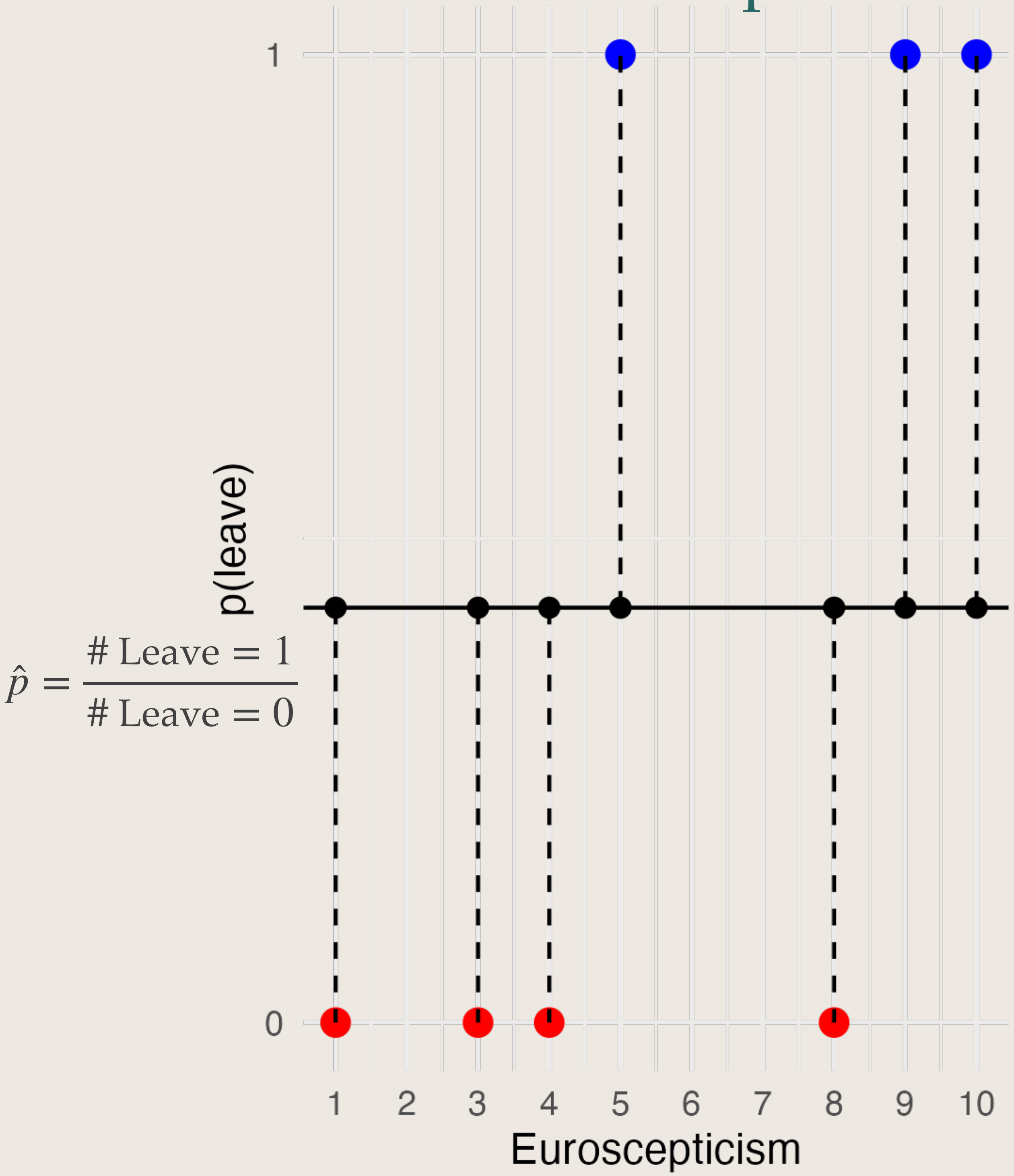


Goodness of fit

Our Model

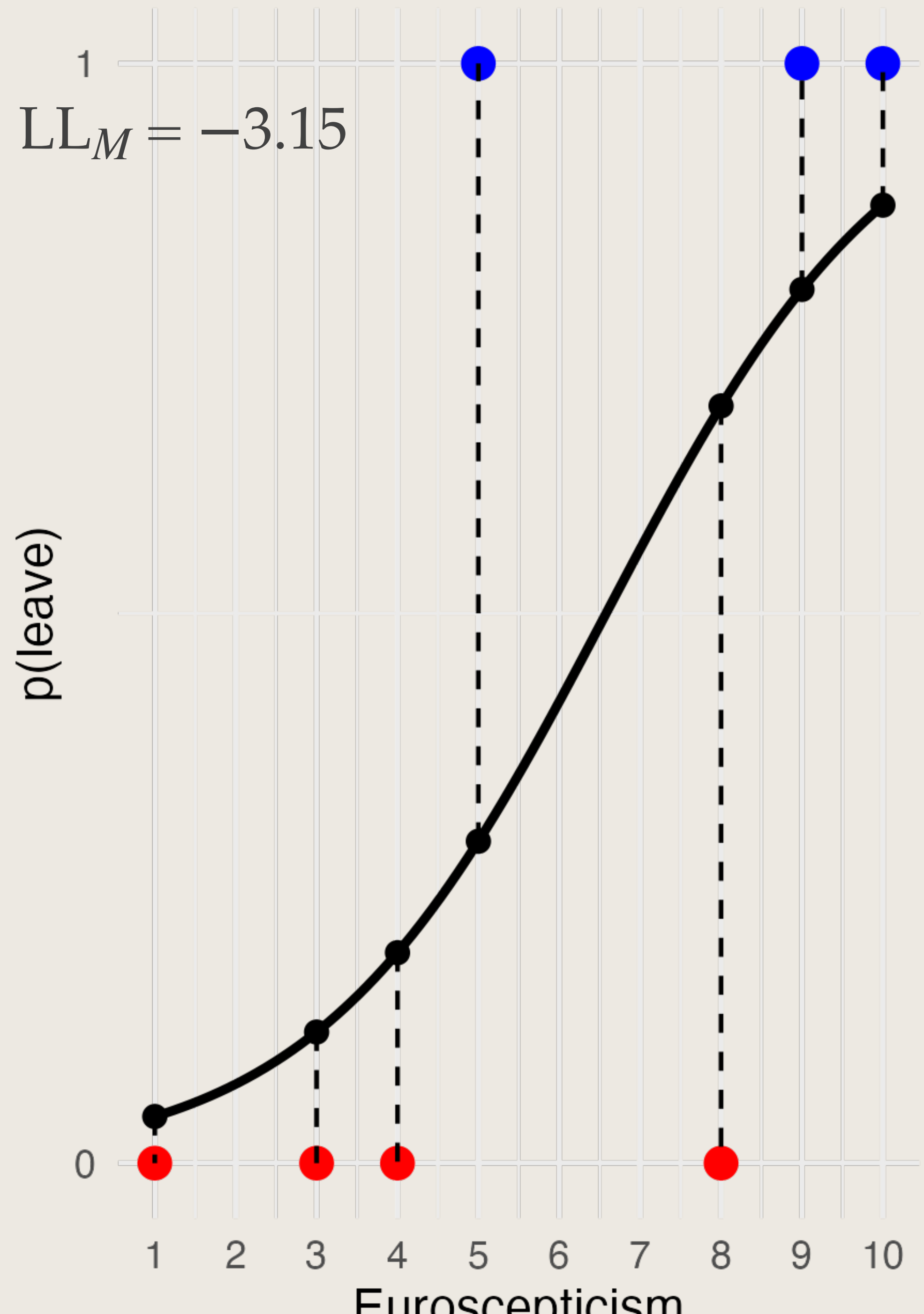


Best Guess without predictors

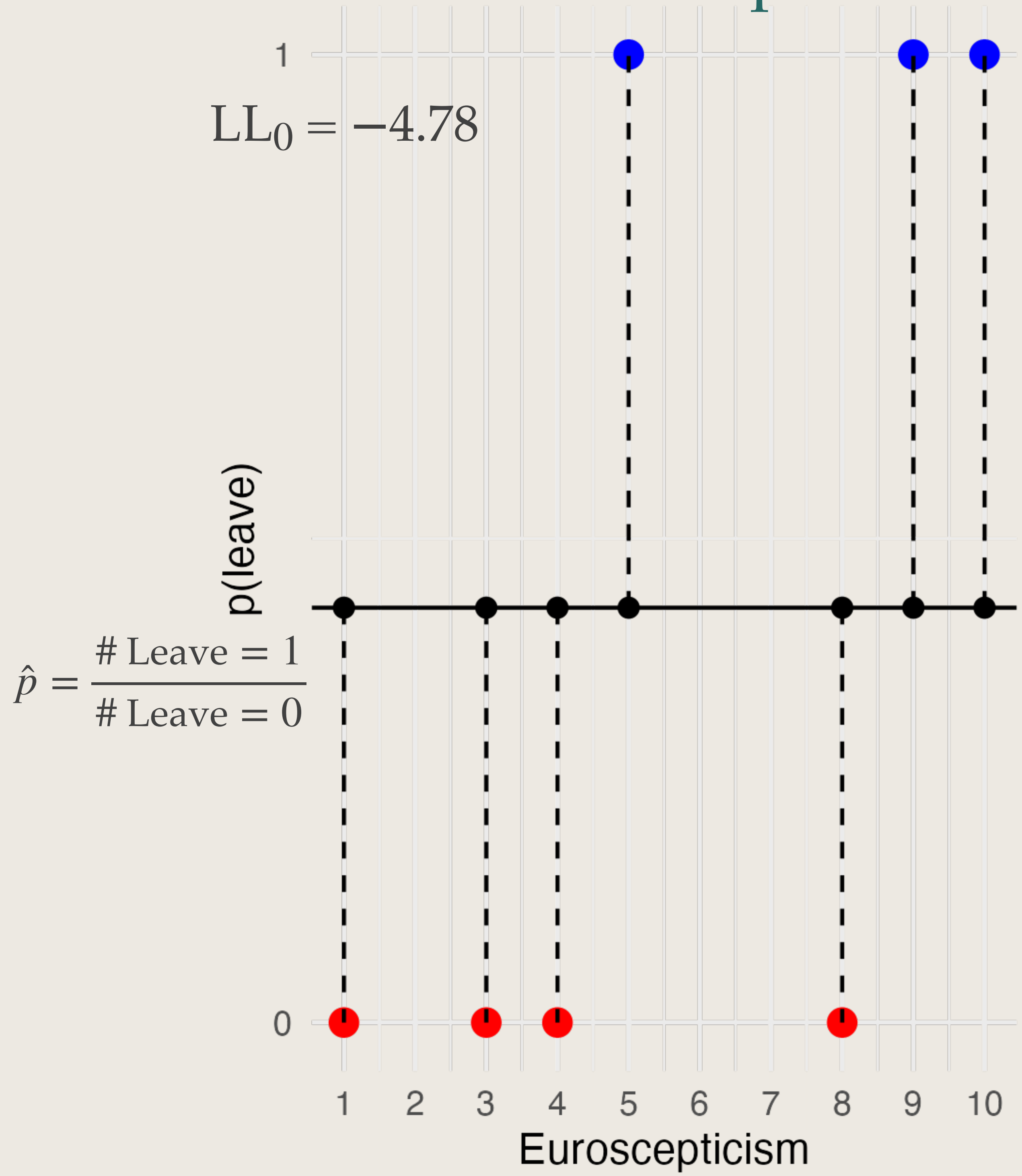


Goodness of fit

Our Model



Best Guess without predictors



Goodness of fit

Goodness of fit

- Simplest option:

Goodness of fit

- Simplest option:

$$\text{McFadden's Pseudo-}R^2\text{: } 1 - \frac{\text{LL}_M}{\text{LL}_0}$$

Goodness of fit

- Simplest option:

$$\text{McFadden's Pseudo-}R^2\text{: } 1 - \frac{\text{LL}_M}{\text{LL}_0}$$

$$\text{Adj. McFadden's Pseudo-}R^2\text{: } 1 - \frac{\text{LL}_M - p - 1}{\text{LL}_0}$$

Goodness of fit

- Simplest option:

$$\text{McFadden's Pseudo-}R^2\text{: } 1 - \frac{\text{LL}_M}{\text{LL}_0}$$

$$\text{Adj. McFadden's Pseudo-}R^2\text{: } 1 - \frac{\text{LL}_M - p - 1}{\text{LL}_0}$$

- Where p is the number of independent variables.

Goodness of fit

- Simplest option:

$$\text{McFadden's Pseudo-}R^2\text{: } 1 - \frac{\mathrm{LL}_M}{\mathrm{LL}_0}$$

$$\text{Adj. McFadden's Pseudo-}R^2\text{: } 1 - \frac{\mathrm{LL}_M - p - 1}{\mathrm{LL}_0}$$

- Where p is the number of independent variables.
- A number of other pseudo-R^2 (Cox&Snell, Nagelkerke, Tjur).

Goodness of fit

- Simplest option:

$$\text{McFadden's Pseudo-}R^2\text{: } 1 - \frac{\text{LL}_M}{\text{LL}_0}$$

$$\text{Adj. McFadden's Pseudo-}R^2\text{: } 1 - \frac{\text{LL}_M - p - 1}{\text{LL}_0}$$

- Where p is the number of independent variables.
- A number of other pseudo-R^2 (Cox&Snell, Nagelkerke, Tjur).
- Any of these goodness-of-fit statistics may be useful for model comparison, but **not worth losing your sleep over**.

Summing Up

Summing Up

* Use logistic regression when you have a binary outcome variable, as you want to restrict predictions to 0-1 interval (i.e. to predict probabilities).

Summing Up

* Use logistic regression when you have a binary outcome variable, as you want to restrict predictions to 0-1 interval (i.e. to predict probabilities).
* Direction and significance of log-odds coefficients are easily interpretable; effect size is not.

Summing Up

- Use logistic regression when you have a binary outcome variable, as you want to restrict predictions to 0-1 interval (i.e. to predict probabilities).
- Direction and significance of log-odds coefficients are easily interpretable; effect size is not.
- Average Marginal Effect is the best way to express how change in X affects the probability that Y is 1.

Summing Up

- Use logistic regression when you have a binary outcome variable, as you want to restrict predictions to 0-1 interval (i.e. to predict probabilities).
- Direction and significance of log-odds coefficients are easily interpretable; effect size is not.
- Average Marginal Effect is the best way to express how change in X affects the probability that Y is 1.
- Use predicted probabilities and AMEs to get a sense of the substantive relationships between variables.

What Next?

What Next?

* Going even further beyond OLS:
  * Other non-linear models (multinomial, Poisson → extensions of logistic regression).
  * ML approaches (Lasso, Ridge, Decision Trees).

What Next?

* Going even further beyond OLS:
  * Other non-linear models (multinomial, Poisson → extensions of logistic regression).
  * ML approaches (Lasso, Ridge, Decision Trees).
* Specific ways of applying our workhorse models:
  * Time series (panel data, survival analysis).
  * Design-based approaches (matching, IV, RDDs).
  * Experiments.

What Next?

- Going even further beyond OLS:
  - Other non-linear models (multinomial, Poisson → extensions of logistic regression).
  - ML approaches (Lasso, Ridge, Decision Trees).
- Specific ways of applying our workhorse models:
  - Time series (panel data, survival analysis).
  - Design-based approaches (matching, IV, RDDs).
  - Experiments.
- Stats for goals other than inference:
  - Measurement, classification, description of complex systems.

What Next?

* Going even further beyond OLS:
  * Other non-linear models (multinomial, Poisson → extensions of logistic regression).
  * ML approaches (Lasso, Ridge, Decision Trees).
* Specific ways of applying our workhorse models:
  * Time series (panel data, survival analysis).
  * Design-based approaches (matching, IV, RDDs).
  * Experiments.
* Stats for goals other than inference:
  * Measurement, classification, description of complex systems.
* Method options are sprawling and changing fast (AI is coming for all of us) — make your methods training fit your research needs, not the other way around.

What Next?

What Next?

- Hilary Term 2024:
  - Causal Inference — for a taste, see Imbens (forthcoming) "Causal Inference in the Social Sciences", *Annual Review of Statistics and Its Application.*
  - Computational Methods (with Rachel!)

What Next?

- Hilary Term 2024:
  - Causal Inference — for a taste, see Imbens (forthcoming) "Causal Inference in the Social Sciences", *Annual Review of Statistics and Its Application.*
  - Computational Methods (with Rachel!)
- Oxford Spring School 2024 (applications now open):
  - Machine Learning
  - Causal Inference (design-based, field experiments)
  - Text Analysis

What Next?

- Hilary Term 2024:
  - Causal Inference — for a taste, see Imbens (forthcoming) "Causal Inference in the Social Sciences", *Annual Review of Statistics and Its Application.*
  - Computational Methods (with Rachel!)
- Oxford Spring School 2024 (applications now open):
  - Machine Learning
  - Causal Inference (design-based, field experiments)
  - Text Analysis
- Trinity Term 4-week courses.

What Next?

- Hilary Term 2024:
  - Causal Inference — for a taste, see Imbens (forthcoming) "Causal Inference in the Social Sciences", *Annual Review of Statistics and Its Application.*
  - Computational Methods (with Rachel!)
- Oxford Spring School 2024 (applications now open):
  - Machine Learning
  - Causal Inference (design-based, field experiments)
  - Text Analysis
- Trinity Term 4-week courses.
- Can't get enough of it? Audit Intermediate Stats next year. Keep an eye out for method courses (ECPR, ICPSR, SICSS, EITM…)

What Next?

- Hilary Term 2024:
  - Causal Inference — for a taste, see Imbens (forthcoming) "Causal Inference in the Social Sciences", *Annual Review of Statistics and Its Application.*
  - Computational Methods (with Rachel!)
- Oxford Spring School 2024 (applications now open):
  - Machine Learning
  - Causal Inference (design-based, field experiments)
  - Text Analysis
- Trinity Term 4-week courses.
- Can't get enough of it? Audit Intermediate Stats next year. Keep an eye out for method courses (ECPR, ICPSR, SICSS, EITM…)
- Long-term investment will involve some self-learning.

Thank you for your kind attention!

Leonardo Carella

leonardo.carella@nuffield.ox.ac.uk