

Models of outcome and choice: The logit model

Silje Synnøve Lyder Hermansen

```
library(dplyr); library(ggplot2)  
theme_set(theme_minimal())
```

Before we start

Before we start

Where are we?

Assumptions of the linear model

Linear models (OLS) rely on two assumptions that are often violated

- ▶ observations are independent and identically distributed (iid)
- ▶ **outcomes are continuous and unbounded** (next 7 weeks)

⇒ *this class: alternative models when these are not satisfied.*

Take 1: A latent variable approach to GLMs

Many outcomes are not continuous

- ▶ **OLS assumes a continuous dependent variable. But many phenomena in the social sciences are not like that.**
 - ▶ Vote choice, civil conflict onset, legislator performance, court rulings, time to compliance, etc.

⇒ *OK. Let's strategize.*

All regressions are linear(ized)

- ▶ **The basic formulation in any regression describes a linear relationship between x_i and y_i :**

$$y_i = \alpha + \beta x_i + \epsilon_i$$

- ▶ When x_i increases with one unit, y_i increases with β units.
- ▶ If that relationship is not linear, we have to make it so:
 - ▶ by recoding the x_i
 - ▶ by recoding the $y_i \rightarrow$ we *linearize*.

A latent variable

- ▶ **A linear(ized) model requires a continuous dependent variable.**
 - ▶ Imagine we are interested in an unobservable variable, z_i , that describes our propensity towards something.
 - ▶ Above a certain threshold (τ) of z_i , observability kicks in and we can see y_i .
 - ▶ The regression coefficients (β) in GLMs describe the $z \sim x$ relationship.

⇒ *The latent variable approach is useful when interpreting the results.*

Example: The binomial model

- ▶ **The logit model is a perfect example:**

$$y_i = \begin{cases} 1 & \text{if } z_i > \tau \\ 0 & \text{if } z_i \leq \tau \end{cases}$$

- ▶ The probability (z_i) of an outcome y_i is continuous.
- ▶ Above a certain probability (τ), we observe a positive outcome ($y_i = 1$).

\Rightarrow *But how do we set the value of τ ?*

From latent variable to discrete outcomes

Statistical theory helps us describe how z_i leads to y_i .

- ▶ **What kind of process generated our data?** → Data Generating Process (DGP)
- ▶ **How can we best describe it?** → choice of *probability distribution* (in GLM)

The three components of GLMs

- ▶ **When fitting the model, we need to make three choices:**
 - ▶ A linear predictor: βx_i .
 - ▶ A probability distribution: they're all in the exponential family.
 - ▶ A recoding strategy.

In R this translates to two additional arguments compared to your usual OLS.

- ▶ A linear predictor: $\rightarrow (y \sim x)$.
- ▶ A probability distribution: $\rightarrow (\text{family} =)$.
- ▶ A recoding strategy $\rightarrow (\text{link} =)$.

The three components of GLMs

- ▶ **In R, this translates to two additional arguments compared to your usual OLS:**
 - ▶ A linear predictor: $\rightarrow (y \sim x)$.
 - ▶ A probability distribution: $\rightarrow (\text{family} =)$
 - ▶ A recoding strategy $\rightarrow (\text{link} =)$.

Example R code for a GLM model

```
mod <- glm(y ~ x,  
           data = data,  
           family = binomial(link = "logit"))
```


Latent variable approach for interpretation

- ▶ The latent variable approach is useful when interpreting results.
- ▶ That's when we map *from* the latent variable *to* the observed outcome.

⇒ *When estimating the model, we have to go the other way round.*

Take 2: Recoding from binary to continuous

How do we get from a binary to a continuous variable?

Data structure

**We can only observe the outcome produced by the latent variable.
There are two data structures for binary data:**

- ▶ classes of observations: e.g.: rats in a cage, coin tosses...
- ▶ case-based: e.g.: legislator votes, Brexit...

Data structure

**We can only observe the outcome produced by the latent variable.
There are two data structures for binary data:**

- ▶ classes of observations: e.g.: rats in a cage, coin tosses... → *the closest to the latent continuous variable.*
- ▶ case-based: e.g.: legislator votes, Brexit...

⇒ *we know the number of successes and trials in a cage/class/stratum.
That's our starting point.*

The binomial distribution: successes and failures

The binomial distribution: successes and failures

How does the binomial distribution map discrete outcomes (0 or 1) to something continuous?

- ▶ let's start with the intercept-only model (no predictors, just a base-line probability)

Let's exemplify with rats

A probability distribution describes the probability of all potential outcomes

- ▶ We kept a 1000 rats in a cage and a number of them died (failure) while others are still alive (success).

⇒ *How can we model this?*

Step 1: describe all potential outcomes

- ▶ Let's consider a series of 1000 potential trials (cages) where we let the successes go from complete failure (success = 0) to complete success (success = 1000)

```
trials <- 1000  
success <- 0:1000  
failure <- trials - success
```

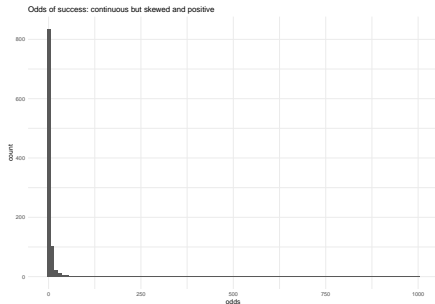
⇒ *We describe all potential outcomes*

Step 2: we calculate the odds

We calculate the odds of surviving in a cage in a 1000 cages

- ▶ compare successes with failures by dividing one by the other

```
odds <- success/failure
```



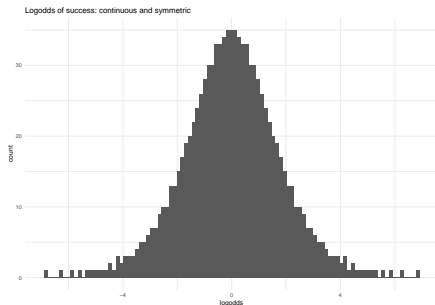
⇒ *A continuous outcome from 0 to + infinity*

Step 3: we log-transform the odds

We logtransform the odds of surviving in a cage in a 1000 cages

- ▶ use the logarithmic transformation: natural logarithm (e) of the odds

```
logodds <- log(odds)
```



⇒ *A continuous, bell-shaped outcome from - to + infinity*

The recoded dependent variable has a linear relationship to x

This, we can run regressions on!

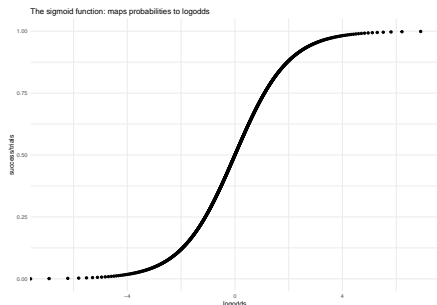
- ▶ the outcome variable in logistic regressions is logodds
- ▶ ... meaning the regression coefficients are reported on that scale

⇒ ... *but they're not easy to understand, so we backtransform when interpreting*

The famous S shape (sigmoid shape)

We can plot the logodds of success against the number of successes or their probability (it's the same).

- ▶ we can go back and forth between logodds and successes/probabilities
- ▶ log-transformation:
 - ▶ forces outcome to be between 0 and 1
 - ▶ residuals are homoscedastic (constant variance)



⇒ *curve “flattens out” when closing up to the 0 or 1 boundary, so relationship is non-linear*

Probability distributions for binary variables

There are two, closely related probability distributions for binary outcomes:

- ▶ The binomial distribution: $B(n, p)$
 - ▶ p is the probability of success tells where on the x-axis (trials) the distribution is placed.
 - ▶ n is the number of trials and defines the precision (spread) of the distribution.
- ▶ The Bernoulli distribution: $Ber(p)$: when we only have only one trial $B(1, p) = Ber(p)$.
- ▶ Data structure: When we have data (covariates) on the event level, we use the case based approach.
 - ▶ y is coded as 0 or 1, R recodes

Why all the fuzz? Why not OLS?

Distributions in OLS and maximum likelihood

- ▶ In OLS: The residuals must be normally distributed (but not the y_i)
- ▶ In ML: The z_i must follow a known probability distribution.

⇒ *This what allows us to translate the latent variable to probable outcomes.*

What happens if I run a linear model on binary outcomes?

- ▶ The model risks predicting out of the possible boundaries
 - ▶ Predictions are wrong.
 - ▶ Regression coefficients are wrong.
 - ▶ Standard errors are wrong.
- ▶ The relationship between x_i and y_i is constant across all values.

⇒ *This last element has a bearing for the interpretation.*

Example

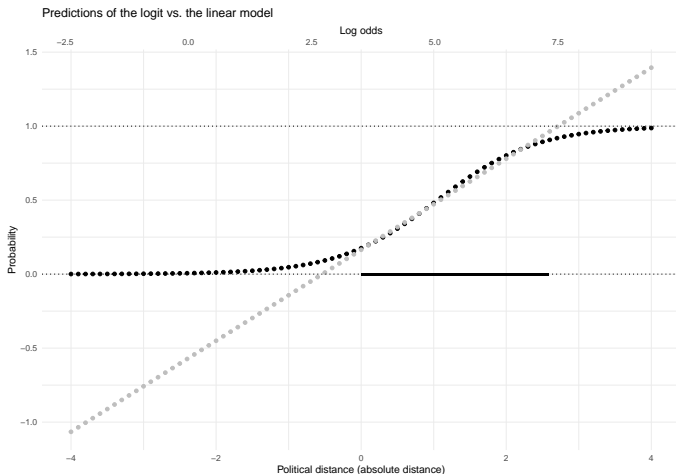
What is the likelihood that a judge at the Court of Justice of the European Union is replaced by another judge at the end of their mandate?

Table 1: Probability of a judge to exit after their mandate ended

	<i>Dependent variable:</i>	
	<i>y</i>	
	<i>OLS</i> (1)	<i>logistic</i> (2)
Political distance between governments	0.308*** (0.069)	1.472*** (0.378)
Constant	0.165*** (0.036)	-1.548*** (0.210)
Observations	251	251
R ²	0.074	
Adjusted R ²	0.070	
Log Likelihood		-138.038
Akaike Inf. Crit.		280.076
Residual Std. Error	0.429 (df = 249)	
F Statistic	19.834*** (df = 1; 249)	
<i>Note:</i>		
* p<0.1; ** p<0.05; *** p<0.01		

Let's back-transform and plot predictions

If we create scenarios for labor cost, we see that at the fringes, the two curves differ.



Interpretation: So... what did I find?

Back and forth: Logistic and logit transformation

Back and forth: Logistic and logit transformation

The logit transformation

When we go from outcomes to latent variable we use the logit transformation.

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

⇒ This what R does when estimating our model

The logistic transformation

When we go from the latent variable to outcomes we use the logistic transformation.

$$\text{logit}^{-1}(\text{logodds}) = \frac{\exp(\text{logodds})}{1 + \exp(\text{logodds})} = \frac{1}{1 + \exp(-\text{logodds})} \quad (2)$$

⇒ This what we do when interpreting our model

My three stages of interpretation

My three stages of interpretation

I go through three stages of interpretation by first setting two scenarios (or more)

- ▶ Marginal effects from regression table : half-way scenario (change in x)
 - ▶ Logodds: check direction and significance (in text).
 - ▶ Odds ratio (for large coefficients) and percentage change (for smaller coefficients).
- ▶ First-difference: full-fledged scenario (all x -s) to make predictions with point estimates (in text)
- ▶ Predictions: a bunch of full-fledged scenarios with uncertainty (graphics).

The regression table: marginal effects

I interpret the regression coefficient itself

- ▶ Change in logodds: check direction and significance.
- ▶ Odds ratio (for large coefficients) and percentage change (for smaller coefficients).

⇒ *A first stab at hypothesis testing.*

The regression table: marginal effects

Now, you try! What statements would you make using the change in logodds, the odds ratio and percentage change? {

Table 2: Judges' likelihood of being replaced (a binomial logit)

	<i>Dependent variable:</i>
	exit
free_economy_diff	1.385*** (0.432)
AgeExit	0.114*** (0.023)
attendance	-0.016** (0.008)
Constant	-8.403*** (1.503)
Observations	236
Log Likelihood	-115.485
Akaike Inf. Crit.	238.970
Note: * p<0.1; ** p<0.05; *** p<0.01	

}

{

The regression table: marginal effects

Typical statements about marginal effects

- ▶ Change in logodds: "When the political distance between judges' appointing and reappointing governments increases, the likelihood of replacing the judge increases."
- ▶ Percentage change (for smaller coefficients; 1.385; change = interquartile range). "The likelihood that a judge is replaced is 71% higher for a judge facing a relatively distant government compared to a colleague tha faces a more aligned government."
- ▶ Multiplicative change (for larger coefficients; 0.114). "The likelihood that a judge exits the court is 1.12 times higher if that distance increased to 1."

⇒ *A first stab at hypothesis testing.*

Predicted values

If you believe the model describes reality appropriately, you can learn more about it by interpreting more thoroughly

- ▶ Odds ratios are notoriously hard to understand.
- ▶ The effect depends on the value of y_i and all the other x s.

⇒ *Interpret the predicted values*

Predicted point estimates (text)

Formulate scenarios using point estimates (in text)

- ▶ Take an all-else-equal approach: Let one x change and keep all others constant (on a typical value).
- ▶ Find the typical representative of two x values and set the other x s accordingly.

⇒ *Which one you use depends on your objective: A theoretical point, assess effect of intervention on groups...*

Predicted values (graphic)

Formulate scenarios using point estimates and put them on speed

- ▶ Predict y values for the entire range of x and plot it.
- ▶ Simulate confidence and plot that too.
- ▶ You can do this for two scenarios.

⇒ *You get a sense of the actual differences in the data.*

Model assessment: How well is reality described?

Model assessment

Model assessments aim to gauge how well we describe the data (i.e. the y).

- ▶ comparison between predicted and observed values (as in OLS).
- ▶ mapping outcomes to the recoded, "latent" variable (GLM).

⇒ *You have a few additional "tricks" to the standard OLS assessment.*

Brier score

Describes the "average size" of the residuals.

$$B_b \equiv \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - y_i)^2 \quad (3)$$

\Rightarrow *Lower scores imply better predictions.*

How well do I discriminate?

How well do I discriminate?

The real question for logits is how well do I distinguish 0s from 1s.

- ▶ what is the value of my cut point (τ)?

⇒ *Several strategies.*

Table comparison

I can set a single cut point.

- ▶ I often use the null-model (i.e. proportion of successes)
 - ▶ then recode all probabilities higher than the cut point to 1 and all below to 0:
- ▶ How often do I predict correctly?
- ▶ on average (proportion of corrects)
- ▶ for each value of the outcome (true/false positives and negatives)

⇒ *I can decide how risk-averse I am in my positive predictions*

The ROC curve

The ROC lets the cut values vary and displays how wrong we are on each side (true positive vs. false positive).

- ▶ A model with good predictions has a curve tending towards the upper left corner.
- ▶ The actual cut value depends on our priorities

⇒ *The graphic is useful in and of itself*

Hosmer-Lemeshow test

Doesn't set the cut point, but bins the data.

- ▶ sorts data from low to high probability
- ▶ slices it up in g number of groups (e.g. by deciles)

⇒ performs a χ^2 test to assess whether the prediction are significantly different from the observations

The separation plot

The separation plot shows how the density of observed "successes" increases as our predicted values increase.

⇒ Another graphic that is useful in and of itself